# Thermal Imaging For Amplifying Human Perception

Von der Fakultät für Informatik, Elektrotechnik und Informationstechnik und dem Stuttgart Research Centre for Simulation Technology (SRC SimTech) der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

vorgelegt von

**Yomna Ali Gamaleldin Abdelrahman**

aus Kairo, Ägypten

| | |
|---|---|
| Erstgutachter: | Prof. Dr. Albrecht Schmidt |
| Zweitgutachter: | Prof. Dr. Frank Vetere |
| Zweitgutachter: | Prof. Dr. Kaisa Väänänen |
| | |
| Tag der mündlichen Prüfung: | 07.12.2018 |

Institut für Visualisierung und Interaktive Systeme
der Universität Stuttgart

2018

# ABSTRACT

Humans throughout history have aspired to own superpowers to extend their control over their surroundings and maximize their benefits. Such superpowers were inspired by advantages of other creatures or by human ingenuity and imagination to extend the motor, perceptual, or cognitive capabilities. The synergy between the machine extended superpower and the human proved to be the optimal combination for highest performance. The work of this thesis focuses on *amplifying the perceptual capabilities of humans via extending their vision*. Visual perception is the ability to see, process and understand stimuli in an environment. Despite its importance, the visible luminance range that is detectable is limited. The perceivable spectrum of the human eye comprises less than one percent of the electromagnetic spectrum. Historically, tools that enhance our vision for the visible aspects of our immediate environment like lenses and glasses were extensively built. Afterwards, the research focus shifted to building tools that support us in seeing what is naturally invisible like X-rays and telescopes.

Thermal imaging is on the brink of being integrated in our daily devices, enabling the shift from specific context usage to mainstream. The first simple products, such as attachable smart-phone cameras, hit the mass market and achieved considerable market attention. However, the number and versatility of study prototypes in the field of Human Computer Interaction is far beyond the available devices on the market. In particular, amplifying our visual perception and making the invisible visible have high potential to change the way we perceive and interact with our environment and surroundings. Thermal imaging allows the unobtrusive sensing and visualizing of the heat map of the environment including both; objects, and human subjects. It is capable of sensing the physiological information of human subjects, detecting actions performed in the environment, and depicting the environment in an amplified form.

In this thesis, we explore how thermal imaging can amplify our visual perception. Employing a user-centered design process, we demonstrate how different thermal properties can be leveraged to make the invisible visible. We focus on amplifying our perception of the environment as well as of the cognitive load. We use a probe-based research approach to systematically investigate the possible potentials and challenges of thermal imaging. We developed ten study probes showing how our perception benefits from thermal imaging and what technical and user requirements thermal imaging poses for amplified perception systems.

We present a contextual inquiry to investigate the understanding and perception of thermal imaging by diverse groups of users. We show that users highlighted that thermal imaging increased their awareness of their environment and changed how they interact with it. On the environment side, we look at implicit amplified perception by presenting the thermal camera feed to the user as well as explicit amplified perception by processing the scene before presenting it to the user. We present three study probes in different forms, namely two head mounted displays as well as hand-held devices to investigate amplified environment perception. In particular, we had a real-life application in the context of firefighters' amplified environment perception.

Beside the amplified perception of the environment, we explore how thermal imaging can be used as a window into our internal states. As thermal imaging remotely measures our body temperature, which reflects information about our internal states and changes in our autonomic nervous system, they allow different physiological states to be sensed unobtrusively. We explore how we can leverage these physiological signals for amplifying our perception of cognitive load. We conduct two studies assessing how we can estimate our cognitive load and classify attention type using these temperature signals.

From the aforementioned study probes, we collected a unique and novel thermal images dataset. With the consent of the involved parties, we release all the collected data sets and the implemented systems as open source. The implemented systems are intended to support developers in deploying and utilizing thermal cameras by providing an easy to use Windows Application Service, to connect and acquire the feed from the thermal camera.

Finally, we distilled a set of design recommendations. These recommendations are grouped into social-based and technology-based recommendations and serve as a basis for designing novel thermal imaging based systems. Throughout our developed and evaluated research and study probes we derive design implications and a conceptual architecture for amplified perception using thermal imaging. We dedicate a discussion to the social implications of using thermal imaging derived through the evaluations of our developed probes. We conclude with a vision of thermal imaging and discuss opportunities of future work.

# ZUSAMMENFASSUNG

Übermenschliche Käfte zu besitzen und damit einen Vorteil gegenüber anderen zu haben, war schon immer ein Wunschtraum der Menschen. Die Inspiration für Superkräfte kommt aus dem menschlichen Einfallsreichtum und der Beobachtung von Fabel- und Lebewesen. Dabei erträumen sich Menschen eine Erweiterung ihrer motorischen, perzeptiven und kognitiven Fähigkeiten. Aus einer technischen Sicht sollen die erweiterten Fähigkeiten durch die Synergie von Mensch und Technik erreicht werden, wobei eine optimale Kombination für höchste Leistungsfähigkeit angestrebt wird.

Der Fokus dieser Arbeit liegt auf der Erweiterung des menschlichen Sehsinns. Visuelle Wahrnehmung ist die Fähigkeit, Stimuli in der Umgebung zu sehen, zu verarbeiten und zu verstehen. Die durch den Menschen wahrnehmbaren Wellenlängen und auch die wahrnehmbare Leuchtdichte sind begrenzt. Dennoch können Menschen zum Beispiel eine 50 Kilometer entfernte Kerze in der Nacht entdecken. Das menschliche Auge erfasst aber lediglich ein Prozent des elektromagnetischen Spektrums. Im Lauf der Geschichte wurden verschiedene optische Geräte und Brillen entwickelt, die den Sehsinn verstärken, um die sichtbaren Aspekte unserer unmittelbaren Umgebung besser wahrnehmen zu können. Der Fokus in der Forschung und Entwicklung verschob sich hin zu Geräten, welche Phänomene sichtbar machen, die zuvor unsichtbar waren, wie zum Beispiel Röntgengeräte und Teleskope.

Wärmebildkameras sind inzwischen weit entwickelt und stehen kurz davor, in verschiedene Geräte integriert zu werden. Die Nutzung wandelt sich von der Spezialanwendung hin zur allmeinen Nutzung in verschiedenen Szenarien. Die ersten einfachen Geräte, beispielsweise Wärmebildkameras für das Smartphone, haben auf dem Massenmarkt eine beachtliche Aufmerksamkeit erreicht. Im Forschungsgebiet der Mensch-Computer-Interaktion gibt es eine zunehmende Anzahl von Prototypen und Studien, welche Wärmebildkameras für interaktive Systeme nutzen. Hierbei birgt gerade die Verstärkung unserer visuellen Wahrnehmung ein großes Veränderungspotenzial. Dadurch dass das Unsichtbare sichtbar wird, verändert sich, wie wir unsere Umgebung wahrnehmen und wie wir mit ihr interagieren. Die Wärmebildgebung erlaubt die Wahrnehmung und Sichtbarmachung der Wärmesignatur einer Umgebung. Dabei werden die Temperaturunterschiede sichtbar, und sowohl Objekte mit bestimmten Eigenschaften als auch Menschen zeichnen sich ab. Es können so auch physiologische Informationen von Menschen erfassen werden. Zusätzlich lassen sich Handlungen in einer Umgebung nachvollziehen, und die Umgebung kann mit zusätzlichen Informationen dargestellt werden.

In dieser Dissertation wird untersucht, wie Wärmebildgebung unsere visuelle Wahrnehmung erweitern kann. Mittels eines nutzerzentrierten Designprozesses wurde untersucht, wie unterschiedliche Wärmeeigenschaften und Temperaturinformationen genutzt werden können, um Unsichtbares sichtbar zu machen. Der Fokus liegt sowohl auf der Wahrnehmung der Umgebung, der Interaktion, wie auch anderer Menschen.

Unter Verwendung eines Prototypen-basierten Forschungsansatzes werden die Möglichkeiten und Herausforderungen der Wärmebilddarstellung systematisch erforscht. In zehn verschiedenen Studien wird dargelegt, wie die menschliche Wahrnehmung von der Wärmebildgebung profitieren kann und welche technischen und Benutzeranforderungen diese mit sich bringt.

Im Kontext realer Anwendungsszenarien wurde das Verständnis von Wärmebildkameras bei verschiedenen Nutzergruppen empirisch erforscht. In der Arbeit werden die Potenziale und Herausforderungen präsentiert. Nutzer haben deutlich gemacht, dass Wärmebildkameras ihre Wahrnehmung für die Umgebung in bestimmten Kontexten verbessert und somit die Interaktion mit ihr verändert haben. Einerseits wurde hierbei untersucht, wie die durch rohe Temperaturinformation implizit erweiterte Darstellung des Videobildes genutzt werden kann. Anderseits wurde betrachtet, wie eine Darstellung explizit durch verarbeitete und interpretierte Informationen aus dem Wärmebild sinnvoll erweitert werden kann. In drei verschiedenen Studien, vor allem mit Head-Mounted-Displays und mobilen Geräten wurde die erweiterte Wahrnehmung der Umgebung untersucht. Hierbei wurde insbesondere der Nutzungskontext Feuerwehr als realer Anwendungsfall untersucht.

Des Weiteren wurde erforscht, wie Wärmebildkameras als Sensor für emotionale und kognitive innere Zustände genutzt werden können. Mit Wärmebildkameras lässt sich die Körpertemperatur genau messen. Damit können Informationen über das Befinden des Nutzers und über Veränderungen im vegetativen Nervensystem erfasst werden. Dies ermöglicht eine unauffällige Wahrnehmung verschiedener physiologischer Zustände. Es wird erforscht, wie diese physiologischen Signale nutzbar gemacht werden können, um die Wahrnehmung für solche Zustände zu erweitern. In zwei Studien wurde gezeigt, dass die kognitive Belastung und die unterschiedlichen Arten von Aufmerksamkeit mit Wärmebildgebung erfasst und gemessen werden können.

Mit den vorgenannten Studien wurde ein neuer Datensatz erhoben und mit dem Einverständnis der Beteiligten öffentlich zugänglich gemacht. Es wurde ebenfalls Software der entwickelten Systeme veröffentlicht, welcher unter anderem Entwickler dabei unterstützt, Wärmebildkameras zu integrieren.

Aus der Forschung werden Designempfehlungen abgeleitet. Diese sind aufgeteilt in soziale und technische Empfehlungen und dienen als Basis für das Design von neuartigen Systemen, welche Wärmebildinformationen nutzen. Auf Basis der Erfahrung aus den verschiedenen entwickelten Systemen und evaluierten Studien werden Auswirkungen analysiert und Empfehlungen für das Design diskutiert. Es wird ebenfalls eine konzeptuelle Architektur für solche Systeme vorgestellt. Abschließend folgt eine Diskussion über die Potenziale der Wärmebildgebung in der zukünftigen Forschung im Bereich der Mensch-Computer-Interaktion.

# PREFACE

This thesis contains work created from 2014 to 2018 at the University of Stuttgart. Since studying of thermal imaging based system requires different types of expertise from different disciplines, this thesis has been done in close collaboration with experts from the University of Stuttgart, project partners within the feuerWeRR project, Microsoft SocialNUI lab and external collaborators. These collaborations resulted in publications which are a core part of this thesis. The contributing authors (i.e. co-authors of papers) are clearly stated at the beginning of each chapter together with the reference to the publication if applicable. To keep the consistency throughout the thesis and to emphasize these collaborations, I use the term "we" instead of "I" when referring to myself.

# Acknowledgments

There can be no words that can express my gratitude for the support I have experienced throughout my PhD from too many people that contributed tremendously to this work. Although it will never be enough nor fair to the amount of guidance, support and love, here is an attempt to acknowledge their massive efforts.

To my supervisor, *Albrecht Schmidt* for supporting me in all ways from signing for an apartment and VISA support letters to professional guidance and advice: Thank you for the support during my Masters and my PhD, and even for my future steps you are already there supporting me. I am sure none of these would have been possible if it wasn't for you. You always had my back literally during my injuries and hospital downtime. Where I will always be thankful for all the things you did and represented, what will always be in my heart and mind is your inspiring, supportive and caring character. I am honored to be one of your team and your number 19. I will never be able to thank you enough.

To my internship host, *Frank Vetere* thank you for hosting me in your SocialNUI, Microsoft research group. Thank you for the meetings we had and interesting discussions over fancy lunch places in Melbourne. Above all, thank you for making my experience a fruitful and unforgettable one that turned out to be the start of so many cool projects.

To *Niels Henze*, thank you for being there. You also had your share of writing and signing random documents for me. Thank you for the hiring Skype call and for mentoring me through both my Masters and PhD studies. It was great working with you and having you in the office next door. I really appreciate and loved our talks and your sarcasm.

To *Pawel W. Woźniak* the frustration melter, thank you very much for all the support, meetings, colored pens and sticky notes and your amazing efforts during GET THE THESIS DONE Skype meetings.

To *Stefan Schneegass*, thank you for your efforts in supporting me and our GUC initiative. It really meant a lot that you supported it in the last minute.

To *Eduardo Velloso*, thank you for the opportunity, guidance and access to awesome people in the SocialNUI group. Working with you was the cherry on top of this thesis. I admit it was a turning point in my PhD and academic career and the start of a long academic collaboration and a lasting friendship, not to mention the amazing BBQs.

To *Joshua Newn*, thank you for being such a great friend, who was there during my down time even though miles away and across time zones. Thank you for making Melbourne feel like home.

To *Tilman Dingler*, thank you for all the work, and all the discussions beyond work. Thank you for all the professional and personal advice and our late deadline night talks. Most importantly, thank you for all the Melbourne pictures and our friendship that survived the distance and time zones.

To *Alireza Sahami Shirazi*, thank you for your time and supervision during my M.Sc. thesis and for the material shopping for our thermal papers. Above all thank you for making my PhD topic possible.

To my mentor, *Slim Abdennadher* for supporting my PhD through sending to HCILab amazing students Dina Hammouda, Abanoub Mimi, Mai Elkomy, Jailan Salah, Elhassan Makled, Sheriene Safwat, Ahmed Zaki, Mohamad Tarek and Karim Labib. Thank you for all the advice, discussions, and proposal writing experiences.

To the very special office mates and friends *Katrin Wolf*, thank you for the nice talks and the walks in different cities around the world. *Alexander Voit*, thank you for nice company and making the office a greener place with your plants.

*Romina Poguntke* thank you for the talks, love and sharing your Brinze cookies and love for kobiba with me.

Thank you *Dominik Weber* for keeping my ice cream intake ongoing.

To Pascal Knierim, thank you for all the rampe meetings and the report translations you had to go through.

To Bastian, thank you Mr. yellow pages for knowing all answers to my random questions.

To *Mariam Hassib*, thank you for pushing the thesis writing move and for being the friend you are.

To *Passant Alagroudy*, thank you for making any travel experience different and unique to the extent sometimes that they are hard to believe or to track. I admit it was fun though.

To the entire HCILab for what it was and will always be one of the greatest teams and labs. Thank you, Markus Funk, Rufat, Jackob, Thomas, Sven, Lars, Huy, Francisco, Patrick, Vali, Miriam, Mauro, and Céline. Thank you for all the great days, BBQs, butter chicken, late night work and for being my family in Stuttgart.

Florain Alt, my future supervisor, yet I had the pleasure to work with you on our CHI award paper. Thank you very much for your support with the random letters for the auslanderbehorde even before starting working together.

To my dearest *Anja*, thank you for our cultural talks, vine leaves making session, your emotional support and holding the group together across Stuttgart and Munich. To the amazing *Murielle*, thank you for your efforts to make me get the DRA correct and your endless support with our paperwork and contracts.

I would also love to thank the people behind the scenes: My family and friends outside the work.

To *Sarah Farid and Ghada Dessouky*, my supporters and emergency contact :), thanks for having my back - literally at times.

To *Gamila ElGarhy*, my sister and extended family. Thank you for everything you are.

To *Sahar Malek, Sarah Sherief, Karim Reda, Mila and Baby Adam*, Stuttgart and my life are brightened with your presence.

To my friends and little sisters, *Lobna Hegazy*, *Mai Elkomy*, *Perihan Sameh* and *Menah Bastawisy*. Where shall I even begin? Thank you for being there and managing to maintain our friendship through distance.

To Tarek ElZomor and Hesham Aboelfadl, words can't describe how grateful and blessed I am for having brothers like you.

To Dr. Ahmed Galal, I am really grateful that our paths have crossed. Thank you for sharing the love of Semsema's shawerma with me, and helping out pushing the last step of publishing this thesis.

To my brothers *Ahmed* and *Abdelrahman*, sisters in-law *Dina* and *Yomna*, thank you for being the best brothers and sisters anyone can wish for. Thank you for bringing joy and happiness into my life through the tiny monsters *Farida*, *Ali*, *Hana* and the tiniest family member yet to come.

To my parents, *Mervat Abdelfatah* and *Ali Zohni* I am speechless, as there is no words can ever express how thankful I am. I know it was hard on you, but thank you for believing in me and supporting me throughout my entire life.

# TABLE OF CONTENTS

**List of Figures** xix

**List of Tables** xxi

**I INTRODUCTION AND BACKGROUND** 1

**1 Introduction** 3

1.1 Research Questions . . . . . . . . . . . . . . . . . . . . . . . . 5

1.2 Research Approach & Methodology . . . . . . . . . . . . . . . . . 5

1.3 Research Contributions . . . . . . . . . . . . . . . . . . . . . . 7

1.4 Research Context . . . . . . . . . . . . . . . . . . . . . . . . . 9

1.5 Thesis Outline . . . . . . . . . . . . . . . . . . . . . . . . . . 10

**2 Background** 15

2.1 Discovery of Infrared Radiation . . . . . . . . . . . . . . . . . . 15

2.2 General Principles . . . . . . . . . . . . . . . . . . . . . . . . 16

2.2.1 Thermal Vs. Infrared Radiation . . . . . . . . . . . . . . 16

2.2.2 Thermal Radiation Modeling . . . . . . . . . . . . . . . . 17

2.2.3 Thermal Properties . . . . . . . . . . . . . . . . . . . . 22

2.3 Thermal Infrared Measurement Systems . . . . . . . . . . . . . . 24

2.4 Thermal Imaging Application History . . . . . . . . . . . . . . 27

2.5 Amplified Visual Perception . . . . . . . . . . . . . . . . . . . 29

2.6 Chapter Summary . . . . . . . . . . . . . . . . . . . . . . . . . 31

**II STUDY PROBES & PROTOTYPES** 33

**3 Understanding Thermal Imaging** 37

3.1 Expert Understanding . . . . . . . . . . . . . . . . . . . . . . . 38

3.1.1 Method . . . . . . . . . . . . . . . . . . . . . . . . 38
3.1.2 Analysis and Results . . . . . . . . . . . . . . . . . . . 41
3.1.3 Survey . . . . . . . . . . . . . . . . . . . . . . . . 41
3.1.4 Focus Group . . . . . . . . . . . . . . . . . . . . . . 42
3.1.5 Discussion and Summary . . . . . . . . . . . . . . . . . 43
3.2 Novice Understanding . . . . . . . . . . . . . . . . . . . . . 45
3.2.1 Mixed-Method Probe Methodology . . . . . . . . . . . 46
3.2.2 Analysis and Results . . . . . . . . . . . . . . . . . . . 50
3.2.3 Discussion . . . . . . . . . . . . . . . . . . . . . . . 58
3.2.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . 60
3.3 Chapter Summary . . . . . . . . . . . . . . . . . . . . . . . 61

**4 Amplified Perception of the Environment 63**
4.1 FeuerWeRR: Amplified Perception for Firefighters . . . . . . . 65
4.1.1 Related Work . . . . . . . . . . . . . . . . . . . . . . 66
4.1.2 Study I: Depth Vs. Thermal Vision Amplificat . . . . . 67
4.1.3 Results and Discussion . . . . . . . . . . . . . . . . . 68
4.1.4 Study II: Hand Held Vs. Head Mounted Display . . . . 69
4.1.5 Results and Discussion . . . . . . . . . . . . . . . . . 69
4.1.6 Summary . . . . . . . . . . . . . . . . . . . . . . . . 70
4.2 TriSight: Amplified Perception for Novice Users . . . . . . . . 71
4.2.1 Related Work . . . . . . . . . . . . . . . . . . . . . . 72
4.2.2 Trisight: A Proof-of-Concept Prototype . . . . . . . . . 73
4.2.3 Evaluation . . . . . . . . . . . . . . . . . . . . . . . 75
4.2.4 Results . . . . . . . . . . . . . . . . . . . . . . . . 77
4.2.5 Discussion and Summary . . . . . . . . . . . . . . . . . 82
4.3 ThermalMirror: Amplified Interaction Space . . . . . . . . . . 84
4.3.1 Related Work . . . . . . . . . . . . . . . . . . . . . . 84
4.3.2 Thermal Reflectivity . . . . . . . . . . . . . . . . . . 87
4.3.3 Evaluation . . . . . . . . . . . . . . . . . . . . . . . 89
4.3.4 Results and Discussion . . . . . . . . . . . . . . . . . 91
4.3.5 Summary . . . . . . . . . . . . . . . . . . . . . . . . 95
4.4 ThermalAttacks: Detecting PINs and Patterns . . . . . . . . . . 96
4.4.1 Related Work . . . . . . . . . . . . . . . . . . . . . . 97
4.4.2 Thermal Attacks . . . . . . . . . . . . . . . . . . . . 98
4.4.3 Understanding thermal attacks . . . . . . . . . . . . . . 99
4.4.4 ThermalAnalyzer . . . . . . . . . . . . . . . . . . . . 101

4.4.5 Evaluation . . . . . 104
4.4.6 Results . . . . . 107
4.4.7 Discussion and Summary . . . . . 114
4.5 VID: Veins Patterns for User Identification . . . . . 116
4.5.1 Related Work . . . . . 117
4.5.2 Vein Identification . . . . . 117
4.5.3 Evaluation . . . . . 119
4.5.4 Results . . . . . 121
4.5.5 Discussion and Summary . . . . . 124

**5 Amplified Perception of Cognitive Load 125**

5.1 CognitiveHeat: Estimating Cognitive Load Level . . . . . 128
5.1.1 Related Work . . . . . 129
5.1.2 Thermal Imaging for Cognitive Load Estimation . . . . . 132
5.1.3 Cognitive load and Facial Temperature . . . . . 133
5.1.4 Temporal Latency of Facial Temperature Change . . . . . 141
5.1.5 Discussion and Summary . . . . . 146
5.2 AttenTCam: Classifying Attention Type . . . . . 148
5.2.1 Related Work . . . . . 150
5.2.2 Attention Type Classification . . . . . 155
5.2.3 Results . . . . . 164
5.2.4 Discussion and Summary . . . . . 173

**III DESIGN IMPLICATIONS & GUIDELINES 181**

**6 Implications and Design Recommendations 183**

6.1 Introductory Phase . . . . . 183
6.2 Application Specific Form Factor . . . . . 184
6.3 Context Awareness and Social Context . . . . . 184
6.4 Privacy and Social Consideration . . . . . 184
6.5 Ethics and Data Collection . . . . . 185
6.6 Technical and Generic Conceptual Architecture . . . . . 186

IV CONCLUSION AND OUTLOOK 189

7 Conclusion and Outlook 191

7.1 Summary of Contribution . . . . . . . . . . . . . . . . . . . . . . 191
7.1.1 Understanding Thermal Imaging . . . . . . . . . . . . . . . 192
7.1.2 Amplified Perception of the Environment . . . . . . . . . 192
7.1.3 Amplified Perception of Cognitive Load . . . . . . . . . . 193
7.1.4 Design Implications and Guidelines . . . . . . . . . . . . 193
7.2 Future Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . 195

V BIBLIOGRAPHY 199

Bibliography 201

# List of Figures


1.1 Outline of this thesis. . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.1 Radiation spectrum . . . . . . . . . . . . . . . . . . . . . . . . . . 16
2.2 Classification of the Infrared Spectrum. . . . . . . . . . . . . . . 17
2.3 Radiation parameters . . . . . . . . . . . . . . . . . . . . . . . . . 20
2.4 Radiation parameters for different surface material . . . . . . . . 21
2.5 Emissivity of different materials [61]. . . . . . . . . . . . . . . . . 22
2.6 Types of reflection . . . . . . . . . . . . . . . . . . . . . . . . . . 23
2.7 Apparent temperature of the glass and the radiating sources . . . 25
2.8 Illustration of how thermal imaging are constructed . . . . . . . . 25
2.9 Illustration of the sensor advancement over the four years. . . . . 27
2.10 Thermal imaging applications evolution. . . . . . . . . . . . . . . 28

3.1 Tunnel fire exercise. . . . . . . . . . . . . . . . . . . . . . . . . . . 39
3.2 Interview during the maintenance duties. . . . . . . . . . . . . . . 40
3.3 Interview setup, in a car parked at the tunnel entrance. . . . . . . 40
3.4 Focus group setup, in the fire station. . . . . . . . . . . . . . . . . 42
3.5 Probe kit: FLIR one, phone, notebook, pen and markers. . . . . . 47
3.6 Traces when touching a surface and a cup. . . . . . . . . . . . . . 53
3.7 Thermal view (right), RGB view (left) . . . . . . . . . . . . . . . 54
3.8 Sink trap blocked(left) and fixed (right). . . . . . . . . . . . . . . 55
3.9 House from outside to identify opened windows or doors. . . . . 56
3.10 Knee injury (left), person's facial temperature (right) . . . . . . . 57
3.11 Traces of footsteps as surveillance technique. . . . . . . . . . . . 58

4.1 Map of the Basement. . . . . . . . . . . . . . . . . . . . . . . . . . 67
4.2 HMD (left) and Hand-held (right) prototype used by firefighters. 69
4.3 TriSight: OculusRift, two RGB cameras and a thermal camera. . 74
4.4 Participants while performing kitchen and office tasks. . . . . . 76
4.5 Images taken with *TriSight* by participants during the study. . . . 77
4.6 Time spent in each view, where the total time is 30 minutes. . . 79
4.7 Amplified Interaction Space . . . . . . . . . . . . . . . . . . . . . 87
4.8 Feature extraction for on surface and in-air gesture. . . . . . . . 89
4.9 Material space for Optris PI160-Based Interaction . . . . . . . . 94
4.10 Recognition pipeline of PINs (top) and patterns (bottom). . . . . 101

4.11 Setup with the thermal camera capturing the phone's screen. . . 105
4.12 The success rate of thermal attack against PINs. . . . . . . . . . . 108
4.13 The mean and standard deviation Levenshtein distances. . . . . 109
4.14 The success rate of thermal attack against patterns. . . . . . . . . 110
4.15 The mean and standard deviation Levenshtein. . . . . . . . . . . . 111
4.16 Approaches for resisting thermal attacks. . . . . . . . . . . . . . . 113
4.17 Veins extraction algorithm. . . . . . . . . . . . . . . . . . . . . . 118
4.18 Study setup indoor and outdoor. . . . . . . . . . . . . . . . . . . . 120
4.19 Thermal view and extracted veins pattern for the four hand poses. 121
4.20 False acceptance and the false rejection rates. . . . . . . . . . . . 122
4.21 False acceptance and the false rejection rates. . . . . . . . . . . . 123
4.22 Potential Use-cases. . . . . . . . . . . . . . . . . . . . . . . . . . 124

5.1 Study setup. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
5.2 Nose and Forehead ROI extraction. . . . . . . . . . . . . . . . . . 135
5.3 The application interface used during the study. . . . . . . . . . 136
5.4 Temperature change between the baseline and the Reading tasks. 138
5.5 Temperature change between the baseline and the Stroop tasks. . 139
5.6 Example for one study sequence. . . . . . . . . . . . . . . . . . . 142
5.7 Nose Temperature Change. . . . . . . . . . . . . . . . . . . . . . 144
5.8 Forehead Temperature Change. . . . . . . . . . . . . . . . . . . . 145
5.9 Forehead-Nose Temperature Change. . . . . . . . . . . . . . . . . 146
5.10 Conditions to stimulate the four different attention types. . . . . 156
5.11 Study setup. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 159
5.12 Area of Interest (AOI). . . . . . . . . . . . . . . . . . . . . . . . 162
5.13 The average cognitive load perceived by the participants. . . . . 164
5.14 Mean temperature change. . . . . . . . . . . . . . . . . . . . . . 167
5.15 Gaze Plots, highlighting the patterns in each task. . . . . . . . . 172
5.16 User Independent Average Confusion Matrices. . . . . . . . . . . 176
5.17 User Dependent Average Confusion Matrices. . . . . . . . . . . . 176
5.18 Average confusion matrices for all classifiers for audio only task. 177

6.1 Thermal camera technical specifications for divers applications. 186

# List of Tables


1.1 Research questions tackled in the course of this thesis . . . . . . 6
1.2 Research prototypes developed in the course of Chapter 3. . . . 11
1.3 Research prototypes developed in the course of Chapter 4. . . . 12
1.4 Research prototypes developed in the course of Chapter 5. . . . 13

3.1 The probe properties. . . . . . . . . . . . . . . . . . . . . . . . . 46
3.2 Demographics of the technology-cultural probe participants. . . 49
3.3 Summary of the usage categories. . . . . . . . . . . . . . . . . . 58

4.1 Mean and standard deviation of the PANAS scores. . . . . . . . 78
4.2 Views used by participants to perform the given task. . . . . . . 80
4.3 The recognition accuracy and the contact point temperature. . . 91
4.4 Success rate and Levenshtein distances for different ages. . . . . 107

5.1 Mean temperature change in the Reading tasks. . . . . . . . . . 137
5.2 Mean temperature change in the Stroop tasks. . . . . . . . . . . 140
5.3 Summary for the onset, saturation and offset in seconds. . . . . 143
5.4 Selected feature set used for classification. . . . . . . . . . . . . 161
5.5 Classification results. . . . . . . . . . . . . . . . . . . . . . . . . 169
5.6 Logistic regression classification performance (all Stimuli). . . . 170
5.7 Logistic regression classifier performance without audio task. . . 171
5.8 Recognition accuracy of each attention type. . . . . . . . . . . . 174

# I

# Introduction and Background

# Chapter 1

# Introduction

*"It is one of the commonest of mistakes to consider that the limit of our power of perception is also the limit of all there is to perceive."*

**– C.W. Leadbeater –**

Humans throughout history have aspired to own superpowers to extend their control over their surroundings and maximize their benefits. Such superpowers were inspired by the advantages of other creatures or by human ingenuity and imagination to extend the motor, perceptual, or cognitive capabilities. For example, inventors used animals as an inspiration and attempted to fly, which led to revolutionizing transportation via airplanes. Similarly, recent research in brain computer interfaces empowered quadriplegic individuals to interact with their environment and enabled humanity to benefit from an individual like Steve Hawking. Another example of human computer interface was enhancing the cognitive abilities of chess players with the support of artificial intelligence machines. In each case, the synergy between the machine (extended superpower) and the human proved to be the optimal combination for highest performance performance.

The work of this thesis focuses on *amplifying the perceptual capabilities of humans via extending their vision*. Visual perception is the ability to see, process and understand stimuli in an environment. Despite its importance, the visible spectrum of the human eye comprises less than 1% of the electromagnetic spectrum. Therefore, historically, tools that enhance our vision for the visible aspects of our immediate environment like lenses and glasses were extensively built. Afterwards, the research focus shifted to building tools that support us in seeing what is naturally invisible like X-rays and telescopes.

Eyeglasses use lenses to correct human visual perception. Lenses can also be used to *amplify* the visual perception. For instance, optical microscopes enable us to perceive objects that would be otherwise too small. Binoculars are devices which let us see things at a greater distance and in greater detail than we would be able to see with the naked eye. Research was conducted on extending our visual perception, for instance Kimber et al. [135] used mirrors to augment user's perception. Others looked into extending the perception via extending the field of view [78, 180, 181]. Recent research explored the visualization of non-visible waves such as wireless traffic [79, 95], where they aimed to extend our perception to include non-visible waves.

Thermal imaging is the capturing of infrared information via a camera. The technology builds upon the research enabling us to see what is naturally invisible. It is highly useful as it captures valuable invisible environmental cues because all objects emit a thermal radiation. For instance, thermal imaging allows visualizing gas/water leakage as they have different temperatures, and can also detect overheating devices. Examples for application domains include firefighting, military and medical applications. During the past decade thermal cameras have witnessed a tremendous reduction in their price, along with great advances in form and usage possibilities, changing it to a commodity device and upgrading its user group. Additionally, it has been integrated in diverse daily devices like mobile phones and drones. However, it is still a growing market and it is not widely adopted yet. Thus, there is a research gap in understanding how thermal cameras are used in daily life beyond specific professional domains to amplify visual perception.

In this thesis, we investigate how thermal imaging could be ubiquitously used to enrich human visual perception. We specifically investigate how typical experts and growing-market novice users perceive and understand the thermal spectrum deploying a mixed method approach including interviews, surveys and cultural-technology probes to gain insights about the users perception of the thermal spectrum. Further, we explore a set of scenarios and use cases using technical prototypes to support the visual perception using thermal imaging. Finally, we discuss a set of guidelines for designing thermal imaging applications.

## 1.1 Research Questions

To understand and investigate how thermal imaging could be used to extend and amplify human perception, two main aspects need to be considered and examined, namely the user perception of the thermal spectrum and the opportunities and enhanced capabilities offered by having this extended perception. Table 1.1 summarizes the research questions, which have driven the research presented in this thesis.

Thermal imaging has drastically decreased in both size and cost, making it commercially available to a wider user group. However, our aim is not only to study thermal imaging due to the "technology push", but rather to examine and investigate how non-traditional imaging, i.e. thermal imaging technology operating in the non-visible spectrum, could be deployed to extend our perception.

Thermal imaging has traditionally been used most by special user groups such as firefighters or technical users rather than non-technical novice users. Therefore, we first focus on the **understanding thermal imaging** by both experts (Research Question (RQ)1) and novice users (RQ2). We further investigate and explore the capabilities of having amplified perception via thermal imaging. Hence, we investigate two main domains in which thermal imaging offers novel approaches or enhances current ones. First, we explore the usage of thermal imaging to **amplify our environment perception** (RQ3). The operation spectrum of thermal cameras enables enhanced sensing of the internal state in a non-obtrusive manner. (RQ4) address how thermal imaging could be deployed to **amplify our perception of the cognitive load**. In the final part of this thesis, we focus on the **implications** of adopting to thermal imaging in our daily lives (RQ5).

## 1.2 Research Approach & Methodology

Although thermal imaging was discovered in the 1800s, with the first thermal camera created in 1929, it has mostly been used by the military and firefighters. However, deploying thermal imaging as a ubiquitous computing tool amplifying human perception and for interactive systems by novice users is a more novel

| RESEARCH QUESTIONS | |
|---|---|
| **UNDERSTANDING THERMAL IMAGING** | |
| **RQ1** | What is the understanding of thermal imaging by expert users? |
| **RQ2** | What is the understanding of thermal imaging by novice users? |
| **AMPLIFIED VISUAL PERCEPTION** | |
| **RQ3** | How can thermal imaging be used to amplify perception of the environment ? |
| **RQ4** | How can thermal imaging be used to amplify perception of cognitive load? |
| **IMPLICATIONS & DESIGN GUIDELINES** | |
| **RQ5** | What are the implications of thermal imaging adoption? |

**Table 1.1:** Research questions tackled in the course of this thesis

area of research. Most of the camera based interactive systems have been using traditional RGB color camera, with a recent interest towards depth cameras e.g. Microsoft Kinect. More recently, most research has focused on designing and evaluating interactive systems and techniques based on imaging technology. Hence, previous research charted design opportunities and constraints, as well as raising concerns of using such technology e.g. privacy issues. Additionally. given that these cameras operate within the perceived spectrum of our eyes, understanding how they operate and capture images is relatively straight forward compared to thermal imaging operating outside the perceived spectrum.

Whereas thermal imaging offers novel opportunities to extend and amplify our perception as well as build novel systems it also raises concerns and constraints. Yet there is no existing work on how novice users perceive this spectrum band. Additionally, the level of their understanding of thermal imaging is not quantified nor clear. Hence, we started by investigating the perception and understanding of thermal imaging by both experts and novice users. Next, we developed a series of research prototypes and probes to answer the research questions and explore the utilization of thermal imaging. All the conducted research and developed prototypes followed a user-centered design approach. Our work provides a coherent understanding of the perception of thermal imaging by users, as well as investigation of potential application domains where thermal imaging offers extended perception via novel approaches or enhancement of existing ones. From our findings, we chart design recommendations for thermal imaging-based interactive systems for developers and researchers. Additionally, we chart the challenges and provide directions for future research.

### Prototypes and Studies

In the course of this thesis, we designed and built ten study probes and prototypes. The developed prototypes ranged from probes to high fidelity fully functioning prototypes. We tested and deployed our prototypes in various setups such as indoor vs. outdoor environments, and lab sessions vs in-home sessions.

We coupled each prototype with several methodologies in the user studies to investigate our research questions thoroughly. We conducted controlled lab, as well as in-situ probe studies. During our studies, we evaluated both the technical and conceptual aspects of the proposed system or prototype. For instance, we evaluated the recognition accuracy of the built system as a quantitative measure. Additionally, we used interviews, focus groups, surveys and probes to collect subjective qualitative data and feedback.

## 1.3 Research Contributions

In this thesis we investigate the usage of thermal imaging to extend and amplify human perception. The contributions of this thesis are fourfold. First, we chart a holistic understanding of how users perceive the thermal spectrum (*C1*), highlighting the potentials as well as challenges for using thermal imaging for amplifying our perception. Additionally, it identifies a set of domains where users can benefit from thermal extended perception. Second, we contribute a set of research explorations in diverse domains and contexts (*C2*). Third, we provide novel datasets of thermal images in diverse contexts (*C3*). Finally, based on the collective findings from the extensive evaluations of the developed systems and prototypes, we present a set of both design and technical recommendations for designing thermal imaging-based systems (*C4*).

**C1: User understanding of thermal imaging** Traditional imaging technology is a mapping to the human visual perception. Hence, users know what to perceive and have an intuitive understanding of the presented images.

Unlike RGB and traditional imaging, thermal imaging operates in the Far Infrared (FIR) beyond the human visual perception. Based on exploration of how users perceive the thermal spectrum through interviews, focus groups and technology-culture probe, we present in chapter 3 a holistic understanding of the users' perception of the thermal spectrum.

**C2: Research Prototypes** The exploration of thermal imaging perception and user understanding revealed several opportunities and capabilities of having amplified perception through thermal imaging. Hence, in the context of this thesis we developed ten research prototypes. The study and research prototypes and their evaluations introduced in this thesis are structured according to their relevant research question. In Chapter 3, the perception of users of thermal imaging is introduced. Two study probes, their evaluations and lessons learnt are presented. The two probes cover different target group: expert and novice. In Chapter 4, amplified environment perception is explored again through five study probes covering hand-held and head-mounted display (HMDs) amplifying perception tools. Finally, in Chapter 5, in three study probes we further investigate amplifying the perception of users' cognitive load. Table 1.4 depicts the developed study probes.

**C3: Open Dataset and Code** The exploration conducted throughout this thesis resulted in a diverse and unique dataset.

**C4: Implications and Design Guidelines** We contribute by exploring the ethics and concerns raised from users' adoption of thermal imaging, evaluating prototypes, and identifying social and technical considerations for designing thermal imaging-based systems. Our explorations provided deep insights into how users perceive thermal spectrum as an amplified perception. Evaluating the built prototypes uncovered the technical specifications of thermal sensors that should be provided in different context and applications. Where perceptual mapping information is easy to specify for traditional RGB imaging technologies, due to the direct mapping to our normal perception. These technical specifications include for instance the thermal sensitivity not perceivable to the users, hence unable to be specified without prior knowledge or experience with thermal imaging. Our exploratory prototypes coupled with reviewing the literature gave us a detailed overview of these technical specifications. Thus, we provide a set of technical and design recommendations which aims to guide future designers, developers and end-user to inform their design decisions.

## 1.4 Research Context

The research covered in this thesis was conducted during my employment between 2014 to 2018 at the University of Stuttgart, in the Human-Computer Interaction group under the supervision of Professor Albrecht Schmidt. The research was inspired by collaborations, publications, and discussions with many experts from different areas.

**FeuerWeRR** The major part of this work was conducted within project, FeuerWeRR [1] with the German Federal Ministry of Education and Research, under the Grant No. 13N13481. In FeuerWeRR, four partner research institutes (Fraunhofer IPA, Fraunhofer IAF, Institute of Signal Processing and System Theory University of Stuttgart, and Institute for Visualization and Interactive Systems University of Stuttgart) set out to enhance Civil Security - Protection and Rescue in complex Work situations through employing a thermal imaging camera with extended reality using radar sensors. By combining technological interventions with firefighters in the first line of operation, this three-year research project (March. 2015 - May. 2018) focused on investigating and enhancing the way firefighters operate and utilize tools, with especial emphasis on thermal imaging and radar sensors.

**University of Stuttgart** Most of the research reported in this thesis was conducted together with colleagues from the University of Stuttgart. Combining the technical knowledge and scientific expertise of the group with my research interests, resulted in a number of publications that are of great importance for this thesis. The collaboration with Pascal Knierim, Stefan Schneegass, Niels Henze, Alireza Sahami Shirazi, Tilman Dingler, Markus Funk, Katrin Wolf, Mariam Hassib, Albrecht Schmidt led to publications within the scope of this thesis (e.g. [3, 6, 8, 11, 12, 39, 69, 70]) and further publications beyond the scope of this thesis (e.g. [1, 7, 16, 71, 130, 179, 213, 275, 276, 277, 278, 279] ).

**External Collaborations** Part of the research covered in this thesis was conducted in cooperation with external colleagues, including the following:

1. **SocialNUI Microsoft Research, Melbourne, Australia**
   The collaboration with Eduardo Velloso. Joshua Newn and Frank Vetere from the SocialNUI lab [2], University of Melbourne led to a series of publications [4, 12, 179].

[1] `https://www.feuerwerr.de/`

[2] `https://socialnui.unimelb.edu.au/`

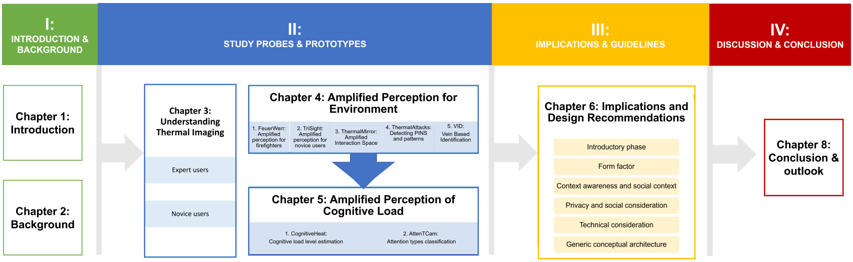



**Figure 1.1:** Outline of this thesis.

2. **German University in Cairo, Cairo, Egypt** A further collaboration, with the German University in Cairo in the context of set of research proposals, included "Serious Games for Education and Sustainable Development", "AffectMyLife" and "Character Computing". The conducted research led to a list of publications in the context of the different research proposal topics(e.g. [16, 17, 63, 71, 130, 163, 211, 212, 213]).

3. **LMU, Munich, Germany** Working with Mohammad Khamis and Florian Alt led to a publication [3], which was recognized by the community and received an honorable mention award for CHI2017. Additionally, part of the covered research in this thesis has been conducted in collaboration with the AMPLIFY ERC project [3] [4, 6, 11, 12].

## 1.5 Thesis Outline

This thesis comprises eight chapters and is divided into four parts, the last two of which contain the bibliography and the appendix. The structure of the thesis closely follows the flow of contributions as depicted in Figure 1.1. The first part introduces the topic of this thesis. The Background part provides an in-depth introduction to thermal imaging and amplified visual perception. This is followed by the two main parts; the study probes and prototypes and implications and design recommendations. In the last part, the conclusion and suggestions for future Work are presented.

[3] http://amp.ubicomp.net

| PROTOTYPE | DESCRIPTION | CHAPTER |
|---|---|---|
| **Understanding Thermal Imaging** | | |
|  | Exploring the expert user perception of thermal imaging by real firefighters. HMDs. *Publication:* [5] | 3 |
|  | **DailyHeat.** In a user study compromising of Cultural-Technology probe and interviews we investigated with novice users in daily home environment *Publication:* [14] | 3 |

**Table 1.2:** Research prototypes developed in the course of Chapter 3.

# Part I: Introduction and Background

**Chapter 1 - Introduction** The first chapter describes the motivation and vision for amplified human perception, states the context of the conducted research, lists the research questions, and summarizes the contribution of our research.

**Chapter 2 - Background** In the second chapter, we introduce the basic foundations and concepts of thermal imaging. Also we present the evolution of the application domains and usage of thermal imaging over the past 20 years.

# Part II: Study Probes and Prototypes

**Chapter 3 - Understanding Thermal Imaging** The chapter introduces two study probes which investigate the prerequisite step in our conducted research of investigating how users both expert (Section 3.1) and novice (Section 3.2) perceive and comprehend thermal imaging.

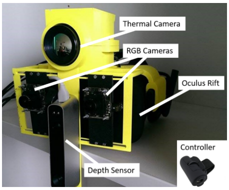
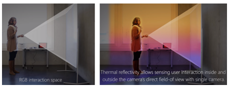
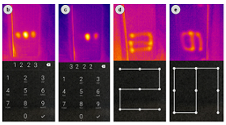
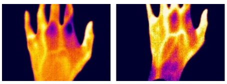
| Prototype | Description | Chapter |
|---|---|---|
| **Amplified Environment Perception** | | |
|  | **FeuerWeRR.** we investigated how they perceive amplified vision via thermal imaging in two forms: (1) Mobile and (2) HMDs imaging. *Publication:* [5] | 4 |
|   | **TriSight.** A HMD prototype compromising of see through using RGB, and amplified vision via thermal camera. We used the amplified vision prototype to evaluate the temporarlity of the amplified vision using set of tasks in three environments; kitchen, office and basement. *Publication:* [15] | 4 |
|  | **ThermalMirrior.** Investigating Thermal Reflection for Amplified Interaction Space *Publication:* [8] | 4 |
|  | **ThermalAttacks** Using the heat traces to infer PINs and Patterns *Publication:* [3] | 4 |
|  | **VID** Identify and authenticate users based on their veins patterns. | 4 |

**Table 1.3:** Research prototypes developed in the course of Chapter 4.

**Chapter 4- Amplified Environment Perception** In this chapter we go along with the first investigation of deploying thermal imaging to amplify environment perception in both context-specific application (cf. Section 4.1) and daily setup (cf. Section 4.2). Through two study probes covering two different contextual scenarios.

**Chapter 5- Amplified Cognitive Load Perception** Different affect states are correlated to our body temperature namely, facial temperature. In this chapter, we explore these insights in the context of affective computing. We aim to explore the usage of thermal imaging to unobtrusively estimate users' internal states to increase humans' awareness of others' internal states. We show that we can estimate the cognitive load level as well as the attention types. Further, we highlight application scenarios of real time, unobtrusive estimation and classification of internal states using thermal imaging.

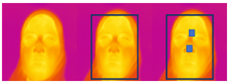
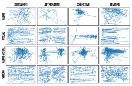
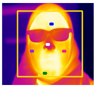
| PROTOTYPE | DESCRIPTION | CHAPTER |
|---|---|---|
| **Amplified Perception of Cognitive Load** | | |
|  | **CognitiveHeat.** Unobtrusive approach to see through users and estimate and quantify their cognitive load level remotely in real time based on their facial temperature. *Publication:* [12]. | 5 |
|   | **AttenTCam.** Classifying users' attention type by augmenting thermal imaging with an eye-tracker. *Publication:* [4]. | 5 |

**Table 1.4:** Research prototypes developed in the course of Chapter 5.

# Part III: Design Implications & Guidelines

**Chapter 6 - Implications and Design Guidelines** In this Chapter, the findings from the study probes are used to derive a set of design recommendations on the user and technical levels. In the second section of this chapter, we synthesize a conceptual architecture for thermal imaging-based systems. The different possibilities and design recommendations are taken into consideration through the different blocks of the architecture. Topics such as privacy, social, and technical considerations are discussed.

# Part IV: Conclusion & Outlook

Finally, part IV summarizes the outcomes of the investigations of the study probes and empirical research conducted, the proposed reference for technical and a conceptual architecture, and a set of design recommendations. Finally, we present our concluding remarks.

**Chapter 8- Conclusion and Outlook** This chapter summarizes the contribution of this thesis. Furthermore, it outlines open questions that still need to be tackled in future developments.

# Chapter 2

# Background

The research presented in this thesis is located in the field of **Amplified Perception using Thermal Imaging**. To understand and utilize thermal imaging, we study the operation features of thermal cameras and the foundation of the thermal spectrum. The main goals of this chapter are first to present basic terms and notions about thermal imaging, and second to present a basic foundation about the thermal infrared spectrum and technologies for thermal cameras. We further discuss how sensing technologies have been used to amplify humans' perception. Related works that directly address the topic and field of the presented prototype and method are further discussed in this and following chapters.

## 2.1 Discovery of Infrared Radiation

The initial discovery of thermal (infrared) radiation was introduced by Sir William Herschel in 1800. In the process of testing the heating properties of different colors of the spectrum, he accidentally discovered infrared radiation. He used the blackened tip of a sensitive mercury thermometer and directed it to a tabletop by having beams of light shine through a glass prism. When he moved the thermometer in the dark area beyond the red end of the spectrum, he noticed an increase in temperature while going down the spectrum from blue to red. A significant increase in temperature was shown as he assessed the temperature of

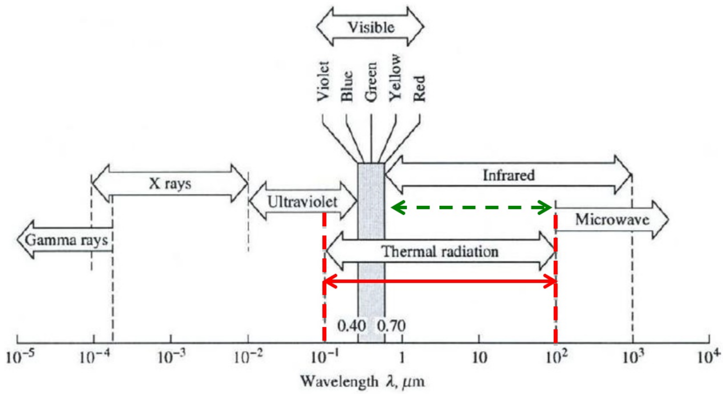



**Figure 2.1:** Radiation spectrum covering the thermal and non-thermal infrared wavelengths and bands only part of the infrared radiation could carry heat, highlighted in dash[5]

the spectrum further beyond the visible spectrum (beyond the red). From here originates the name infrared (below the red).

## 2.2 General Principles

Throughout the thesis we use the terms and notions explained in this chapter. This chapter is written for the reader who is unfamiliar with, or has limited knowledge about thermal imaging.

### 2.2.1 Thermal Vs. Infrared Radiation

There is a common inaccuracy when referring to thermal and infrared radiation as if they were interchangeable. Both terms are commonly cross used. However, the term ***Thermal Radiation*** refers to heat transferred by electromagnetic radiation due to an object's temperature, whereas, ***Infrared Radiation*** is a sub-band of the

[5] http://www.castool.com

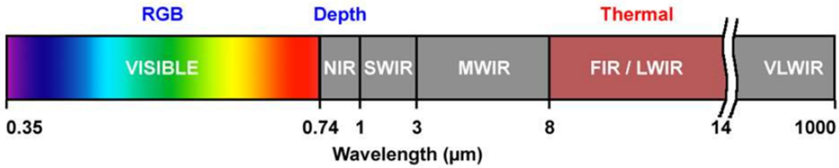



**Figure 2.2:** Classification of the Infrared Spectrum.

electromagnetic spectrum as depicted in Figure 2.2. The source of misconception comes from the fact that the infrared electromagnetic waves can also transfer heat, hence it is sometimes referred to as heat radiation. However, not the entire infrared spectrum can transfer heat. Figure 2.2 depicts the thermal vs. the non-thermal infrared radiation bands. Where the thermal band operating between 7.5 to 13 $\mu$m is utilized in thermal imaging and heat visualization, non-thermal infrared is used in other technologies such as depth cameras and infrared TV remote controls. Figure 2.2 shows the infrared sub-bands including Near Infrared (NIR), Short Wave Infrared (SWIR), Mid Wave Infrared (MWIR), FIR and Very Long Wave Infrared (VLWIR). In this thesis, the focus is on the FIR operating between 7.5 to 13 $\mu$m. Hence, we will use the following terms throughout the thesis:

- IR (infrared): The near, short wave, mid wave, and far infrared energy
- Thermal Radiation: The mid and far infrared energy
- Thermal Imaging: A device using mid or far infrared energy for imaging

### 2.2.2 Thermal Radiation Modeling

In this chapter we present only the basic thermal radiation properties and modeling required to understand the upcoming chapters of this thesis. Further in depth explanation about thermal radiation and the underlying theories and properties can be found elsewhere in the literature, for example [232, 113].

Thermal radiation is generated by the thermal motion of an object's particles. Any object with a temperature above absolute zero [6] emits thermal radiation. Thermal

[6] The lower limit of the thermodynamic temperature scale at $-273.15°$on the Celsius scale.

radiation is emitted due to the object's temperature; the higher the temperature the more the object radiates. Max Planck, Josef Stefan, Ludwig Boltzmann, Wilhelm Wien and Gustav Kirchhoff precisely defined and established qualitative and quantitative correlations for describing infrared energy. This thermal energy and radiation is modeled by the black body radiation model as explained below.

## Black Body Model

The black body is an ideal body that absorbs all electromagnetic radiation applied on its surface irrespective of the angle and frequency of the radiation. In other words, a black body is an abstracted physical body which absorbs all incoming radiation. In other words it does not have any reflective nor transmissive properties. The name black body originates from the fact that it absorbs all visible electromagnetic energy and is perceived as black in color. It changes to red then orange to white hot when it heats up, as stated in the Wein Displacement Law [136].

When the black body is in a thermal equilibrium state (i.e. at a constant temperature) it emits radiation according only to its temperature. It is an ideal absorber (i.e. absorbs all applied radiation) and ideal emitter of radiation over all wavelengths (i.e. emits the maximum energy at all wavelengths) 2.1.

$$\alpha = \varepsilon = 1 \tag{2.1}$$

Where, $\alpha$ :absorption

$\varepsilon$ :emissivity

The main properties of this model were defined by a set of laws introduced in the 1900 by Planck [197], Wien, Kirchhoff, Stefan and Boltzmann [205] as follows:

1. The spectrum of the emitted radiation $M_{\lambda S}$ is described by Planck's law, which presents the basic correlation for non-contact temperature measurements: It describes the spectral specific radiation $M_{\lambda S}$ of the black body into the half space depending on its temperature T and the wavelength Î».
2. The quantity of emitted energy is given by Stefan-Boltzmann's law.
3. The frequency likelihood of the radiation is defined by Wien's displacement law.

**Planck's law** Planck's law defines a distribution that represents the energy level of the radiation at each wavelength (spectral specific radiation $M_\lambda$). The representation is in terms of temperature T and wavelength $\lambda$.

$$M_\lambda = \frac{C_1}{\lambda^5} \frac{1}{e^{C_2/\lambda T} - 1} \tag{2.2}$$

| | | |
|---|---|---|
| Where, | $C_1$ | $3.74 \cdot 10^{-16}$ W m$^2$ |
| | $C_2$ | $1.44 \cdot 10^{-2}$ K m |

This distribution has peak values at a certain wavelength. This peak value has shorter wavelengths for higher temperatures; in other words, the wavelength of the peak of the object's radiation is inversely proportional to its absolute temperature [136].

**Wien's displacement law** As illustrated, the wavelength at the peak decreases as the temperature increases. Wien's displacement law denotes the wavelength at peak and states that it is inversely proportional to the absolute temperature. By differentiating Planck's formula a representation for Wien's displacement law could be derived as shown in equation 2.3.

$$\text{Max}\lambda = \frac{2898}{T} \mu m \cdot K \tag{2.3}$$

**Stefan-Boltzmann law** Finally the Stefan-Boltzmann law defines the spectral radiation intensity at all wavelength values. By integrating these values the emitted radiation of the full body is computed. As shown in equation 2.4 the entire emitted radiation of a black body increases proportionally to the fourth power of its absolute temperature [136].

$$\text{Emitted radiation} \propto \text{Absolute temperature} \tag{2.4}$$

## Gray/Real Body

As mentioned earlier the black body is an ideal model, hence few bodies in the real world act like this; most radiat less emission than the black body given the same temperature. Nevertheless, the black body model is very useful reference to describe the radiation properties, including absorption/emissivity, reflectivity

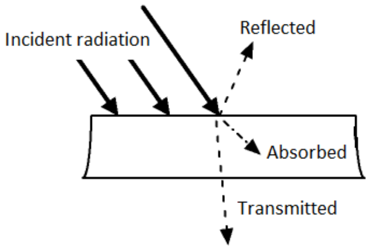



**Figure 2.3:** Radiation parameters

and transmissiveness. The total percentage of the radiation energy could be represented by figure 2.3 and parameters equation 2.5.

$$\alpha + \rho + \tau = 1 \tag{2.5}$$

Where, $\alpha$ : Absorption
$\rho$ : Reflectivity
$\tau$ : Transmissivity

### Absorption and Emissivity

Absorption refers to the amount of thermal energy absorbed by the body. Since the absorbed energy increases the body's temperature it is retransmitted (emitted) to reach equilibrium. Hence:

$$Absorption = Emissivity \tag{2.6}$$

$$\varepsilon + \rho + \tau = 1 \tag{2.7}$$

Where, $\varepsilon$ : Emissivity

*Emissivity*
Emissivity refers to the emitted radiation leaving the surface. It is dependent on several factors including the temperature of the emitting surface, the material the surface is made of, and the properties of the surface (for example roughness). The emission from a surface is distributed among the wavelengths in the thermal band.

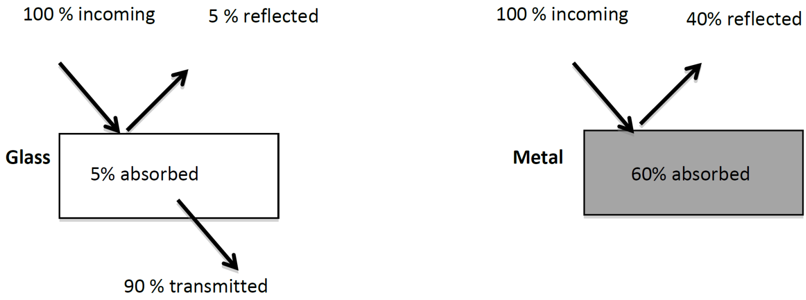



**Figure 2.4:** Radiation parameters for different surface material

The emissivity values range from zero to one. Since the black body is an ideal emitter it has the emissivity value of one. This property measures the radiation ability of an object relative to that of the black body at the same temperature (i.e. the ability of an object to mimic the behavior of the black body).

#### Transmissivity

Transmissivity refers to the ability of an object to pass through thermal energy. Since most objects cannot transmit energy this property can be ignored, as presented in equation 2.8 for simplicity.

$$\varepsilon + \rho = 1 \tag{2.8}$$

#### Reflectivity

Finally, the property of reflectivity is concerned with the degree of reflected energy off a body. From equation 2.8 we can deduce an inversely proportional relation between the reflectivity and emissivity of an object (cf. Figure 2.5).

The reflectivity of an object's surface depends not only on the direction of the incident radiation but also on the direction of the reflection. Surfaces are assumed to reflect in two manners: specular and diffuse. In specular reflection, the angle of reflection equals the angle of the radiation beam. For diffuse reflection the radiation is reflected equally in all directions regardless of the incident radiation's direction. The reflectance of a surface depends on its roughness and the wavelength of radiation strikes [31].

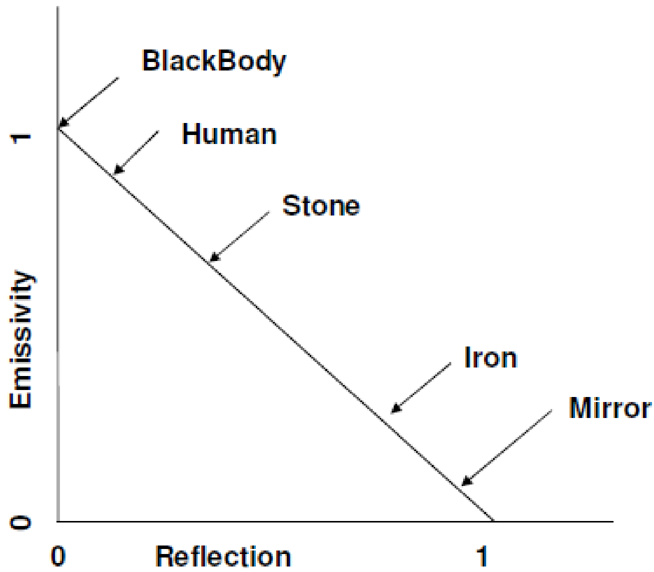



**Figure 2.5:** Emissivity of different materials [61].

### 2.2.3 Thermal Properties

The thermal spectrum exhibits a set of phenomena that differs from the visible light. In this section, we give a brief overview of the main phenomena that have been used in this thesis.

#### Thermal Contact Conductance

One of the main thermal phenomena is the heat transfer from one object to another. For instance when we touch a surface, heat transfers from the users' hand to surfaces they interact with, leaving traces behind that can be analyzed. This relies on the surface's material property know as **thermal contact conductance** [49], which refers to the conductivity of heat between two objects (surfaces) that are in contact.

As described earlier, according to the black body model [113], any object above absolute zero (e.g. surrounding objects in our environment) emits thermal

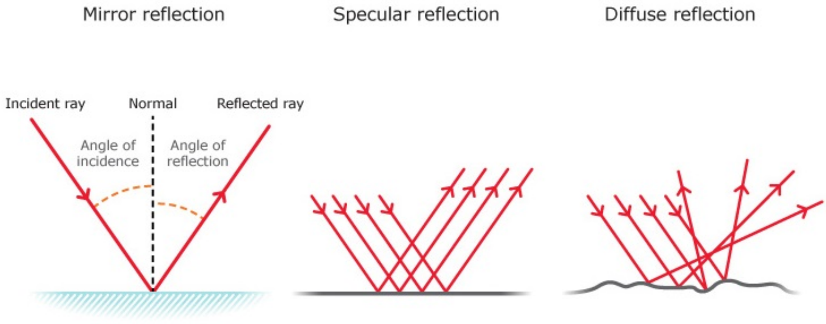



**Figure 2.6:** Types of reflection

radiation. This radiation is absorbed, reflected, and transmitted. However, for fully opaque surfaces the transmitted portion is discarded [81]. This limits the portions of interest to the reflected and absorbed radiation. Hence, thermal radiation can be presented as in $Thermal\ reflectivity + Thermal\ absorptivity = 1$. Thermal absorptivity depends on the temperature of the source that generates the radiation, and is independent of the surface's temperature.

As soon as an object contacts the surface of another object, thermal radiation is transmitted and absorbed by the surface, causing a temperature change. This leads to heat traces accumulating on the surface. Ray [207] built a well-established model to compute the the transferred heat, in other words, the temperature at the contact point of the two bodies.

$$T_{contact} = \frac{b_{object1}T_{object1} + b_{object2}T_{object2}}{b_{object1} + b_{object2}} \tag{2.9}$$

$$b = \sqrt{K.P.C} \tag{2.10}$$

$T_{contact}$ depends on the temperature of the contact points ($T_{object1}$ and $T_{Object2}$) as well as their **thermal penetration coefficient** ($b$). It is the amount of thermal energy penetrated and absorbed by the surface. The $b$ is defined in Equation 4.4. It is composed of the product of *thermal conductivity (K)*, *thermal density (P)*, and *specific heat capacity (C)* [190].

### Thermal Spectral Reflectivity

In the case of the existence of a radiating source, the apparent temperature of the surrounding surfaces is affected as it influences the emitted and reflected radiation from the body. The apparent temperature is directly proportional to the temperature of the radiating source. In the manner that, if the radiating source has a lower temperature than the real body, the apparent temperature will be lower than the real one. Correspondingly, if the radiating source has a higher temperature than the real value the apparent temperature will show a higher temperature [7]. Figure 2.7, illustrates the effect of the radiating source on the apparent temperature computed by the thermal camera. The glass is the body of interest, where three cups of different temperatures are placed on it as radiating sources; one with the same temperature as the glass (room temperature), another with a higher temperature (35°C) and the third with a lower temperature (12°C). As revealed in figure 2.7 the glass's apparent temperature is represented by the middle cup as it has the same temperature as the glass (it is not affecting the apparent temperature), unlike the others: For the colder cup a lower apparent glass temperature and for the hotter cup a higher apparent temperature are detected. The direction of the reflection is defined by the reflective surface properties namely the surface roughness. Surfaces could reflect radiations either in specular or diffuse manner. Surface roughness correlates with how surface reflectivity occurs, on the grounds that any surface with roughness less than one eighth of the radiation wavelength exhibits a specular reflection.

Reflectivity of surfaces differs in the visible and infrared spectra. For instance, aluminum is not reflective in the visible spectrum although it is in the FIR spectrum. Hence, we cannot depend on the behavior of surfaces in the color spectrum and have to measure their roughness to specify the reflectivity in the FIR spectrum, and measure or view them in the thermal spectrum.

## 2.3 Thermal Infrared Measurement Systems

Thermal images are formed from the emitted thermal radiation. Most infrared cameras are built on the same model, as illustrated in Figure 2.8 [8], where the measurement chain integrates the following elements [283]:

[7] http://www.optotherm.com/emiss-physics.htm

[8] https://www.fluke.com/en-us/learn/best-practices/measurement-basics/thermography/how-infrared-cameras-work

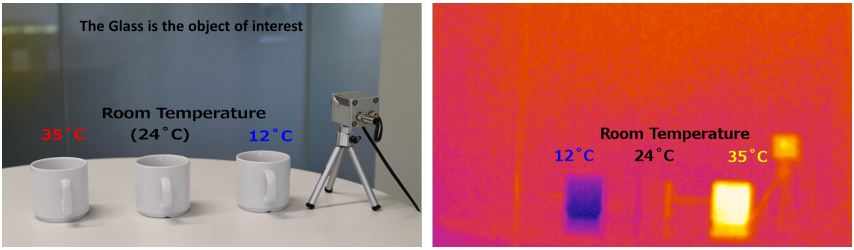



**Figure 2.7:** Apparent temperature of the glass and the radiating sources

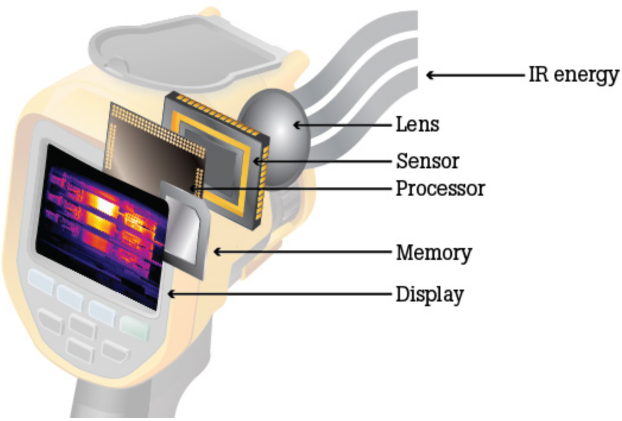



**Figure 2.8:** Illustration of how thermal imaging are constructed

- A special lens focuses the infrared emissions of all the objects in the field of view of the camera. The focused infrared emission is than scanned by the sensor which is also known as the infrared-detector.
- The sensor creates a detailed pixel-based temperature pattern called a thermogram. This thermogram is then transformed unto electric impulses, which are sent to the processor and translated into data for display.
- The processor sends the information to the display, where it is mapped using various colors based on the temperature i.e. intensity of the infrared emission.

Thermal imaging measurement systems prices range from a few hundred dollars up to hundred of thousands. The main factor behind the price difference is the type of the sensor, i.e. the infrared detector. There are two main types of detectors:

- Un-cooled is the common, relatively cheap and consumer-grade thermal imaging device. The sensor operates at room temperature with built-in battery, and operates in real time.
- Cooled on the other hand is more expensive and operates at zero degrees Celsius. The sensor is placed in a cooling container that cools down and maintains the operating temperature. This type of thermal imaging operates with very high resolution and sensitivity due to the cooling mechanism. However, cooled systems are much more expensive and susceptible to wear and tear than uncooled ones.

Over the past decades uncooled thermal imaging devices have advanced drastically. As shown in Figure 2.9 [9], thermal sensors, namely uncooled infrared detectors, have decrease in cost and size. This makes them more widely available and affordable. In the course of this thesis, all the developed prototypes utilized uncooled thermal cameras in different forms including FLIR One, Optris PI160, and Optris PI640 [10].

[9] `https://www.flir.com`

[10] `https://www.Optris.com/`


**Figure 2.9:** Illustration of the sensor advancement over the four years.

## 2.4 Thermal Imaging Application History

*This section is part of the following planned publication:*

- Y. Abdelrahman and A. Schmidt. The history of thermal imaging in human computer interaction *TOCHI*

In the last twenty years, thermal imaging has become more portable and commercially available, thus drawing the attention of Human-Computer Interaction (HCI) researchers to explore it in greater depth as sensing technology. Researchers have explored a wide range of application domains. We briefly summarize the most dominant of these. We are not the first to attempt to classify work using thermal imaging: Ioannou et al. [117] classify thermal imaging usage in Psychophysiology, identifying its potentials and limits. As stated earlier, the early usage of thermal imaging was limited to medical and military use; however, usage has now flourished to cover diverse domains as depicted in Figure 2.10. For instance, Larson et al. used thermal imaging as novel interaction sensing

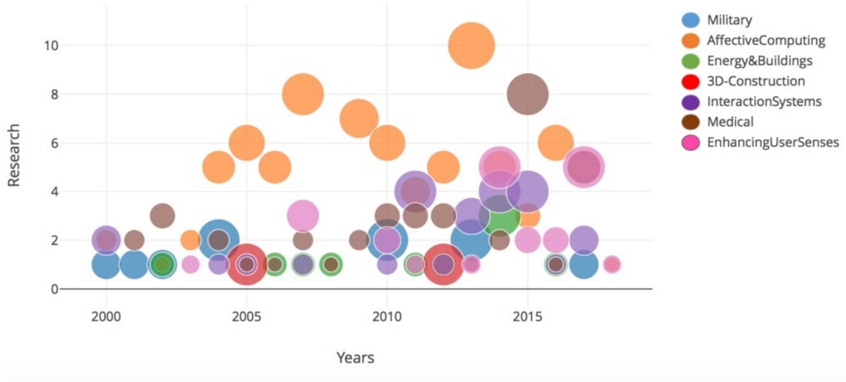



**Figure 2.10:** Thermal imaging applications evolution.

technology based on the heat traces detectable by thermal cameras [150]. Wong et al. used thermal imaging and proposed algorithms to detect a person's head positions and parameter ratios such as the height and the width of objects [281]. Such algorithm were not only used for anti-theft surveillance systems, but also applied in nursing homes for monitoring Alzheimer patients and preventing them from leaving the nursing home without being attended to. The same approach is used for faint detection [280]. Pham et al. proposed a system for posture analysis in a dense crowd of people to detect irregular behavior using thermal imaging [195].

Despite the above-mentioned advances and the significant potential thermal imaging creates both from a research as well as from a commercial perspective, a holistic view of the research on thermal imaging based systems in the past decades is missing as of today. Hence, in the context of this thesis, we aim to provide a holistic view of the conducted research. We identify seven major applications areas of thermal imaging research, namely 1) military, 2) medicine, 3) energy and building, 4) affective computing, 5) 3D construction, 6) interactive systems, and most recent, 7) enhanced sensing. We then summarize existing research in these different areas over the last twenty years.

Our review focused on the research conducted under the umbrella of HCI rather than general research on thermal imaging in different domains for instance in medical field. Hence, we exclusively used Association for Computing Machinery (ACM) and Institute of Electrical and Electronics Engineers (IEEE) digital libraries as our source using the keywords *Thermal Imaging, FIR and HCI*.

Another review and classification iteration was conducted on the cited papers in each retrieved paper to ensure that we covered the related research conducted over the past twenty years. We used the number of papers published in this year as an indicator of the research conducted. Although medical and military research is not usually published or accessible to the public, these were still the leading fields of research in the early 2000s. Taking a deeper look into thermal imaging-related publications, we find that the application domains move more towards the human-centric. For instance, the domains evolved to include interactive systems and focus on daily setups rather than special use cases. Yet most of the conducted research in daily setups as of today misses the holistic understanding of thermal imaging and there are no clear design guidelines for thermal imaging application design.

In summary, to novice users, thermal infrared imaging has been associated with complex and military based application. However, HCI researchers have showed interest in exploring thermal imaging in diverse domains rather than focusing on the military use cases. With the general advancement and miniaturization of sensors, namely thermal imaging, their usage has migrated from limited military-specific to daily and HCI applications. This opens up the opportunity of deployment and interchange between the HCI domain and thermal imaging as sensing technology. In the context of this thesis, we focus on the bidirectional added value for both fields. Where thermal imaging would act as a novel sensing technology in HCI context and domain; on the other hand, HCI would provide a user-centered understanding of thermal imaging. Furthermore, HCI researchers can study and provide human-centric guidelines for designing thermal imaging-based systems, hence bridging the gap between the complex application specific thermal imaging usage as well as the understanding of users' perception and requirements. This understanding will yield into the development of novel thermal imaging-based systems in diverse domains governed by HCI field. Additionally, this would pave the road to deploying thermal imaging in amplifying our perception.

## 2.5 Amplified Visual Perception

Our eyes can only perceive light falling in the visible light spectrum, so the information about our surrounding we can perceive through the electromagnetic spectrum is naturally limited. Evolution shaped humans' visual system making it very well suited to enable their survival in nature. Other animals had other requirements and evolution consequently enabled them to perceive other parts of the electromagnetic spectrum. However, modern society has radically changed the environment we live in, placing additional demands on our visual capabilities.

Almost throughout the entire human history, tools have been built to extend the limitations of human visual perception. Even before the invention of optics, simple tubes were used to reduce stray light. The development of optics, which enables a broad range of tools, can be traced back at least to ancient Egypt where polished crystals were used as lenses. Over the last three millennia, a large number of tools, including glasses, telescopes, and microscopes, have been built to extend human visual perception.

Eyeglasses use lenses to correct human visual perception. Lenses can also be used to *amplify* the visual perception. For instance, optical microscopes enable us to perceive objects that would otherwise be too small. Binoculars are devices that let us see things at a greater distance and in greater detail than would be possible with our own bare eyes. Research has also been conducted on extending our visual perception in additional ways, for instance Kimber et al. [135] used mirrors to augment user's perception. Others looked into extending the perception via extending the field of view [78, 180, 181]. Recent research has explored the visualization of non-visible waves such as wireless traffic [79, 95], which aims to extend our perception to include non-visible waves, in this work we build upon this research and investigate the thermal spectrum as a means to amplify human perception through the non-visible waves.

Tools that enable sensing beyond the visual spectrum have a much shorter history. Infrared rays were only discovered in 1800. In recent decades, novel tools were developed that use cameras which capture images beyond the visual spectrum. For example, thermal cameras are widely used in certain domains such as health care or firefighting. Until recently, they were used primarily by firefighters to make potential hazards such as gas leaks or hot objects visible and to search for persons or sources of heat, as thermal radiation is less scattered by smoke. In addition, researchers explored potential applications for thermal imaging. However, most of these applications address specific domains ranging from augmenting the vision in special scenarios [6, 11] to interactive surfaces [210].

## 2.6 Chapter Summary

In this chapter, we presented an overview of the relevant domains to the work presented in this thesis. Starting with the basic foundations of thermal imaging and the infrared spectrum, we present its operation properties and modeling (cf. Section 2.2). Additionally, in Section 2.3 we give a brief description on the thermal infrared measuring systems and their evolution. In Section 2.4 we present thermal imaging application history. Finally, in Section 2.5, we highlight current approaches to amplify our visual perception.

In Chapter 1 we set the scope and the context of the presented thesis as *thermal imaging and amplified visual perception* from a human centered, HCI perspective. However, this thesis covers diverse domains, which are introduced later in the relevant chapters and sections, illustrating their relevance and connection to our work.

# II

# Study Probes & Prototypes

# OUTLINE

In the following part of the thesis we investigate the understanding of thermal imaging based on two different target groups, namely experienced firefighters and novice everyday users. Thereby, we present work using interviews with firefighters and in-home probes with novice users (cf. Chapter 3). Both approaches aimed to understand how diverse users perceive and understand thermal imaging. Our findings highlight a set of potential usages of thermal imaging as means of extending our perception of the environment (cf. Chapter 4), by utilizing the visualization of the heat distribution and thermal map of the scene, i.e. amplifying our environmental perception to cover the thermal infrared. In addition to amplified environment perception, we present work on how thermal imaging can be used to reveal insights about the internal state of the surrounding people (cf. Chapter 5). Since body temperature are closely connected to users' states, different affect states can be inferred [116, 117]. Additionally, certain temperature patterns act as biometrics and they are uniquely associated to each individual. The uniqueness of a person's biometrics has been used elsewhere to identify and authenticate users [50, 123]. These patterns yield promising biometrics which can be used as implicit user identification (cf., Chapter 5).

This part includes the following three chapters:

- **Chapter 3 - Understanding Thermal Imaging**. Thermal Imaging operates in the non-visible to humans part of the spectrum, hence a crucial prior step before presenting it to users is the investigation of their level of thermal imaging understanding and perception. In this chapter, we present a series of qualitative and quantitative investigations of thermal imaging understanding. Our investigations show that users easily learn and understand the presented thermal information via self exploration, yet highlight the need of an application layer that informs users about the presented information rather than the raw thermal data.

- **Chapter 4 - Amplified Perception of the Environment**. Our visual perception allows us to only perceive our environment in the visible spectrum, which barely comprises 1% of the entire spectrum. Researchers have built tools and optical solutions to increase users' perception and awareness of the environment. In this chapter, we present five prototypes that enable users to perceive the environment in the thermal spectrum. In an evaluation, we show that users intuitively used the thermal view when the visible light perception failed to present sufficient information about the environment.

- **Chapter 5 - Amplified Perception of Cognitive Load**. Different affect states are correlated to our body temperature, namely facial temperature. In this chapter, we explore these insights in the context of affective computing, to discover how thermal imaging can unobtrusively estimate a person's internal state and impart this awareness to others. We show that we can estimate the cognitive load level as well as the attention types. Further, we highlight application scenarios of real time, unobtrusive estimation and classification of internal states using thermal imaging.

# Chapter 3

# Understanding Thermal Imaging

*"Heightened perception is the goal: becoming more aware of how you see, not just what you see."*

**– Michael Kimmelman–**

After setting up and introducing the research question in Chapter 1 and presenting the foundation of thermal imaging and research conducted in amplifying human visual perception in Chapter 2. In this Chapter, we start by investigating the perception and understanding of the thermal spectrum as it is a crucial component for designing thermal imaging based interactive systems to amplify users' perception.

A prerequisite step in our conducted research is to investigate how users both expert and novice perceive and comprehend thermal imaging. Hence, we first set out to understand how thermal images are perceived by the firefighters, given that they are the expert group of using such a tool. We further present an approach to investigate the perception of thermal imaging by non-technical novice users.

Therefore, we conducted qualitative investigations in form of interviews and focus groups with the firefighters to build a deep understanding of their thermal imaging perception. Additionally, we deployed a mix-method technique in the form of a technology-cultural probe to explore how novice users might perceive thermal imaging.

Throughout two study probes targeting two user groups and profiles, we explored various facets of thermal imaging perception from expert and novice perspective. In Section 3.1, we explored the expert understanding of thermal imaging. We extended our exploration to investigate the novice perspective as presented in Section 3.2.

The research questions we address in the presented chapter are:

- **RQ1**: What is the expert user understanding of thermal imaging ?
- **RQ2**: What is the novice user understanding of thermal imaging ?

## 3.1 Expert Understanding

Firefighting is an activity that is highly dependent on the implicit knowledge and perception of the personnel involved. Firefighters build personal knowledge through training and service experience. The ordinary mode of operation of firefighters includes entering buildings full of smoke with limited visibility. Their task is to explore the environment and build a shared understanding of the situation, and decide on the next steps. The availability of tools used e.g. thermal cameras provides information about the environment and offers opportunities for better coordination and improved safety. Hence, we first set out to investigate their perception of thermal imaging being one of the tools used. Additionally, we targeted the firefighters group as they are considered to be typical/exert users of thermal cameras

### 3.1.1 Method

The main objective of our research was to get insights concerning the thermal imaging perception, experiences and practices of firefighters, especially those on the first line of operation. In order to understand how firefighters perceive and

**Figure 3.1:** Tunnel fire exercise.

use imaging technology in their work, we conducted a series of empirical studies with firefighters from Reutlingen, Germany, in the context of the FeuerWeRR [11] project. In the following, we present the studies upon which we build our insights.

## Initial contextual inquiry

We were invited to the fire station in Reutlingen and were introduced to the firefighting process. During a tour of the fire station, the team leader explained the typical process and tools used by the firefighters in the first line of intervention. Additionally, he reported some main challenges regarding the current thermal imaging visualization e.g. no depth information and thermal reflection from mirrors and reflective surfaces. Next, we attended a tunnel fire exercise (Figure 3.1), where we observed the firefighters in operation. We monitored how they use different tools, with special focus on how they use thermal cameras. The entire exercise was video recorded.

## Interviews

To gain deeper insights into how firefighters in the first line of intervention perceive current thermal imaging technology and its visualization, we conducted semi-structured interviews. To that end, we visited the fire station and queried participants while they were performing maintenance duties(cf. Figure 3.2). We also conducted interviews during the tunnel exercise (cf. Figure 3.3). The semi-structured interviews lasted an average of 13 minutes ($SD = 4.10$). We interviewed six firefighters, two of whom were volunteers and four were professionals. The participants were aged 24–45 ($M = 35.17$, $SD = 6.54$). All interviewees were male. All the sessions were video and audio recorded upon receiving consent from the firefighters.

[11] `https://www.feuerwerr.de/`

**Figure 3.2:** Interview during the maintenance duties.

**Figure 3.3:** Interview setup, in a car parked at the tunnel entrance.

We first asked about demographic data and information about their profession (years of experience, position in the team and if it was a volunteer job). Then, we asked about their experience with firefighting and the tools they were using. We focused on the challenges and opportunities of the used tools. We also investigated how they perceived and used thermal cameras in their existing form and visualization. Moreover, to acquire insights from real fire incidents rather than the training context where the interviews where conducted, we asked them to recall specific real life cases.

## 3.1.2 Analysis and Results

We used thematic analysis to analyze the content of the interviews. Our findings highlighted that all the firefighters showed a high level of understanding of thermal imaging on the theoretical level rather than the technical level, as they all responded with the exact definition when asked about thermal imaging. They knew the color mapping of the temperature, and the challenges faced when using thermal imaging such as thermal reflection and lack of depth information. However, surprisingly they showed a lack of a deeper practical understanding. In other words, they knew the information it provided, however they did not posses extensive hands-on experience.

> It's a grayscale image with the dark and light gray color for cold and hot temperatures respectively.*(Firefighters definition)*

Having analyzed the recorded data, we found that tools were shared and not assigned to a single person e.g. each group had a single camera that was held by one firefighter and could be shared with others. This might have led to the lack of hands-on experience, as thermal cameras are a shared tool rather than a personal one. One of the main challenges reported by the firefighters was the lack of depth information in thermal vision. The ability of reporting this reflects their concrete understanding of thermal camera capabilities. Indeed, thermal cameras provide temperature visualization for the field of view of the camera and no depth information.

## 3.1.3 Survey

Based on the interviews, we designed an online survey to explore visualization preferences in a larger and more diverse population sample. Our online survey further investigated different environment visualizations using thermal cameras considering different color mapping as well as distance visualization. To that end, we designed low-fidelity prototypes of different visualizations for thermal data. The survey contained questions about whether the firefighters understood the visualization. Further, we explored if the firefighters would use the presented visualization during emergencies. Finally, we asked if they would change the current visualization to the proposed visualization. The survey was distributed to the firefighters in Germany. We collected $n = 52$ responses (aged 18–58, $M = 33.36$, $SD = 9.99$).

**Figure 3.4:** Focus group setup, in the fire station.

The preferred visualization was one with distance representation in the form of a circle with changing diameter depicting different distances. Firefighters highlighted the need for quick representation with minimal effort to comprehend the distance, hence they discarded using numbers to represent distance. Further, the survey responses reflected the willingness of the firefighters to adopt different thermal images with distance visualization while maintaining the familiar grayscale of thermal images.

### 3.1.4 Focus Group

Next, we conducted a focus group to explore, confirm and contextualize the outcomes of the survey. We focused on how the top rated visualizations in the survey were perceived by the firefighters. The focus group lasted for 90 minutes, and it was conducted in the meeting room of the fire station in Reutlingen (Figure 3.4). We provided colored sticky notes to document their ideas and feedback. The structure of the focus group consisted of an introduction followed by open discussion about the presented visualizations.

In the focus group firefighters interacted with the prototypes and discussed possible usage scenarios for the devices. We had five male participants (aged 21–37, $M = 28.6$, $SD = 6.22$). The focus group outcomes stressed the willingness of the firefighters to adopt new visualizations and new tools. They highlighted they would use the additional information as hints and still trust and rely on their perception.

Concerning the form factor they noted advantages and disadvantages for both forms. For instance, having a head-mounted display would allow them to use both hands for the operation, but hand-held prototypes allowed on-demand operation of the camera. The firefighters rated their experience after using two different prototypes, showing that the head mounted display prototype outperformed the hand-held one. Additionally, they appreciated the distance visualization and quickly grasped the concept with no prior training.

### 3.1.5 Discussion and Summary

In this work we aimed to reveal the understanding of the firefighters of the working spectrum of thermal imaging. Surprisingly, our findings showed that the firefighters have a theoretical rather than practical understanding of thermal imaging, as thermal imaging is a shared tool and not used on daily basis by all firefighters. Additionally our investigation reveled the design constraints and opportunities for fighters from an HCI perspective. Our extensive design inquiry into designing new artifacts for firefighters involved a variety of methods that allowed us to build an understanding of everyday work at the fire station. Here, we present our key observations and implicit lessons learned during our inquiry.

**Trust Dynamics** We observed that trust played a central role in the firefighters' experience of work. Even unprompted, our participants would often reflect on trust and establish clear needs in terms of trust. Firefighters put extensive trust not only in the skills of their immediate colleagues, but also in the way their work was organized and the quality of their training. Most importantly from an interaction design perspective, trust in the quality of professional tools and attire was also very strong. This indicates that designers working with firefighters should skillfully navigate trust dynamics and conduct extensive user studies to ensure that new interactive artifacts evoke trust.

**Team Dynamics** Contrary to our initial expectations, we observed a flexible management structure in firefighter units. While commanding officers were strictly in control of high-level decisions, teams assigned to the same fire engine had a rather flat structure. Firefighters would operate a rotating roster where they performed different duties on different days. This also enabled balanced skill development as teams often consisted of novice and experienced personnel. Future tools for firefighters should reflect these flat structures and enable swift resource and experience sharing.

**Procedures** A high reliance on trained procedures was also prevalent in the firefighter work that we observed. We suspect this was highly related to trust in one's training, which strengthened the ability to put oneself in danger for a greater cause. While experienced firefighters saw procedures and handbooks as necessary and providing a reasonable reference, less seasoned participants were determined to adhere to procedures. Our work shows that engaging with firefighter training and action handbooks (usually present in fire engines) is necessary for building improved interactive tools for firefighting.

**Practicalities** Finally, our work highlights the specific practical aspects of working with users such as active firefighters. The nature of their work implies that they may be available, but may also suddenly quit any research activities. For safety reasons, we were only allowed to see the essence of their work, for example how they would save people in dangerous conditions only during an exercise. Further, as the social environment of the fire station is heavily based on trust, we also needed to build the firefighters' trust in our skills and research agenda. Consequently, we recommend that designers working with firefighters invest time in building trust, make good use of firefighter exercises and allow for enough time spent at the fire station to account for the unpredictable character of firefighting work.

## 3.2 Novice Understanding

*This section is based on the following publication under review:*

- Y. Abdelrahman, P. wozniak, N. Henze, and A. Schmidt. Exploring the potential of thermal imaging usage at home by novice users *CHI'19*

After exploring how typical and expert users perceive and comprehend thermal imaging, we aimed to extend our exploration to reach a wider group, namely novice non-technical users. Thermal imaging operates in a different spectrum than what is visually perceived by our eyes, making its perception and understanding not that straightforward for humans as most do not understand how infrared cameras operate, especially if they have not encountered thermal imaging before. This contrasts with RGB cameras and images which confer a direct mapping to human vision capabilities.

In this study probe we focused on perceiving and experiencing the technology, and motivate the participants to engage with the thermal imaging to gain an overview of their understanding of the technology in hand. We used interviews to acquire a deeper insight into the users' lives. We conducted the exploration in the form of a technology-cultural probe, and as we aimed to gather results of high environmental validity, we decided to use commercially available imaging technology and deployed state-of-the-art consumer-grade thermal cameras in households for 10 days. To fully explore the understanding and usage of thermal imaging, we prepared a probe kit including diaries, note cards, and a commercial mobile phone with an attachable thermal camera FLIR one as shown in Figure 3.5.

Our research methods build on previous research in using cultural design probes and in-home experience [114, 162, 217, 225, 257]. Culture probes were first introduced in 1999 by Gaver et al. [88]. They are a method to elicit creative thinking by asking users to document their ideas and actively involving them in the design process [258], and have been beneficial in exploring technologies before they become widely available and are to be used by non-experts. On the other hand, Schmidt [216, 217] highlights the importance of functional prototypes. In this probe, we focus on thermal imaging, hence we combine technology-cultural probes by using a functioning commercial device as well as diaries and notes for our cultural probe deployed in participants' homes.

| PROPERTY | JUSTIFICATION |
|---|---|
| Functionality | Our technology probe is a simple Android application, with a single main function of capturing thermal pictures and other easily accessible functions e.g. capturing videos. |
| Flexibility | We added the sticky notes and diaries to offer an open-ended experience to encourage users to reinterpret the captured images. |
| Usability | Technology probes usually are not concerned with usability, hence we used the same kit with no iteration on the usability based on the participants' feedback. |
| Logging | Technology probes should help the researchers visualize and analyze the usage of the participants and further discuss innovative ideas of the technology. Our probe complies with this factor, as we logged the captured data (photos, videos and notes) as well as the time of use (timestamps). |
| Design Phase | Our probe was conducted in the early stage of the design process. |

**Table 3.1:** The probe properties.

### 3.2.1 Mixed-Method Probe Methodology

According to Hutchinson et al. [114], technology probes should have the following properties: functionality, flexibility, usability, logging and design phase. Our probe complies with these factors as presented in table 3.1. We chose an easy to use form of the technology to encourage the participants to engage and interact with it, as well as to explore how participants perceive and use it on their own. We combined data from the technology-cultural probe with semi-structured interviews before and after the experience.

**Figure 3.5:** Probe kit: FLIR one, phone, notebook, pen and markers.

## Probe Kit

Our participants were given a probe kit like the one shown in Figure 3.5 containing the following:

1. Diary for on-the-go ideation,
2. 50 empty cards to draw ideas on,
3. Pen and colored markers,
4. Motorola Moto G Smartphone to attach the camera to it,
5. FLIR One for Android [12] thermal camera and the camera's dedicated application [13].

The probe kit enabled participants to easily use a thermal camera. The thermal camera is easily attachable to the smart phone. Using the FLIR app, they can view and record thermal photos or videos. Previous work aimed to prompt ideation via identifying a set of activities to be performed [225]. However, in our work

[12] `http://www.flir.com/flirone/android/`

[13] `https://play.google.com/store/apps/details?id=com.flir.flirone&hl=en`

we aimed to openly explore the understanding and usage of thermal imaging in everyday settings. Hence, we did not use any prompts for activities.

## Study Timeline

We designed our 10 day study as follows:

**Pre-Interview**
Before starting the study we visited the participants' households. We collected their consent for taking part, and conducted a pre-interview to collect their demographics as well as their experience with thermal imaging and whether they had taken part in any previous technology or culture probes.

**Setting up the probe**
Participants were asked to place their probe kit in a location in the home that would be accessible to everyone. They were asked to document their ideas using the cards and diary, and record thermal photos and videos. Additionally, each participating household was presented with a brief introduction to the application, the capabilities of thermal imaging as well as the common properties and features of thermal radiation. We presented examples for the basic features of thermal imaging including:

1. Viewing Thermal Information: by using the camera to view cups with different temperatures.
2. Thermal transfer visualization: by showing the view from the thermal camera while a person touches a cup.
3. Thermal reflection: by placing the camera in front of a reflective surface.

**Semi-Structured Post Interviews**
After 10 days, we revisited the participants' households and gathered data including photos, videos and notes. We conducted a semi-structured interview to gain deeper insights into their experience and how they perceived and used the thermal camera. All household members were invited to the interview. We browsed the captured ideas and photos and asked them to provide explanations and details of the situation/use-case in which they captured the thermal photos and videos. Participants were encouraged to discuss the use-cases with the researchers as well as among themselves to reflect on their understanding. The main goal of the interviews was an in depth exploration of the level of understanding of thermal imaging and the appropriateness of thermal cameras in everyday settings.

| Home | Participant | Gender | Age | Occupation |
|---|---|---|---|---|
| 1 | 1 | Female | 29 | Dermatologist |
| | 2 | Male | 33 | Radiologist |
| | 3 | Male | 62 | Doctor |
| | 4 | Female | 2 | |
| 2 | 1 | Female | 25 | Dentist |
| | 2 | Male | 30 | Dentist |
| | 3 | Female | 60 | Doctor 60 |
| | 4 | Male | 1 | |
| 3 | 1 | Female | 27 | PhD student |
| | 2 | Female | 27 | Manager |
| 4 | 1 | Female | 29 | Housewife |
| | 2 | Male | 27 | Student |
| | 3 | Male | 28 | Engineer |
| | 4 | Female | 2 | |
| 5 | 1 | Female | 21 | Environmentalist |
| | 2 | Male | 22 | Student |
| | 3 | Male | 27 | Guitarist |
| 6 | 1 | Male | 33 | Researcher |
| | 2 | Female | 19 | student |
| 7 | 1 | Male | 27 | Student |
| 8 | 1 | Female | 22 | Housewife |
| | 2 | Male | 23 | Student |
| 9 | 1 | Male | 27 | Student |
| | 2 | Male | 21 | Student |
| 10 | 1 | Female | 21 | Student |
| | 2 | Female | 23 | Student |

**Table 3.2:** Demographics of the technology-cultural probe participants.

## Participants

We conducted the study with 10 households over 10 days. Nine households had at least two individuals. We had 26 (23 adult) participants (11 female) with an age range of 1-62 (mean:28.57, SD:16.97). Most participants held at least a Bachelor's degree. Three of the households had one child and two households had a dog. Only two participants were familiar with thermal imaging, and seven were familiar with the term 'thermal cameras' but had never seen one before. We used numbers for the houses as well as for the participants in each house for referral purposes. H13 means, household 1 and individual number 3 in the household. Guardian consent was acquired for the participation of minors in the study.

### 3.2.2 Analysis and Results

The interviews were audio recorded and transcribed for analysis. We used thematic analysis [37] to analyze the content of the interviews. We analyzed 270 captured photos and 563 recorded interview minutes. Two coders coded 50% of the corpus independently using nVivo [14]. Afterwards, they met to assess differences and construct the final coding tree. The rest of the corpus was coded by a single coder. Through iterative discussion, the final themes emerged from the coded quotations. We present the four main themes reflecting thermal imaging in the home: *Understanding of thermal spectrum, Experiencing Extended Visual Perception, Potential for domestication, Social Implications and When and Why People use Thermal cameras?*

### Understanding of Thermal Imaging

Surprisingly participants showed a clear understanding of the technology by exploration; there was no special training or instructions needed to understand and use thermal cameras.

> It took me couple of views to know the color meaning and then it was easy to understand the hot and cold objects. *(H41)*

Additionally, they reported that their perception easily adapted to the new spectrum. Furthermore, they expressed how their understanding of the thermal images and views created sort of learning experience about the temperature cues as reported by *H61*.

[14] `https://www.qsrinternational.com/nvivo/home`

> It feels like my eyes and mind adapt to the temperature information. *(H12)*

> Surprisingly even when I don't have the camera anymore I am using the hints I already got from the camera, it like how your eyes adapts to this kind of information. *(H61)*

## Experiencing Extended Visual Perception

Participants reported a positive experience regarding accessing and experiencing thermal imaging. They were interested in what thermal cameras allow them to see and sense. Participants recognized the benefits and enhancement in perceiving the environment through thermal cameras:

> It was really cool to have a tool that can compensate what you can't see in hand. *(H32)*

One participant perceived thermal vision as a way to broaden their perception of the world:

> Being able to see the thermal view makes it a totally different experience and makes me feel superior. It's kind of a new dimension of perception. *(H11)*

Others were positively surprised by the ability to notice more details of the environment and thus being able to avoid danger more effectively. One participant stated that they felt safer:

> It felt as I can see the whole surrounding even in the dark which made me feel a bit safe. *(H22)*

> It was actually interesting it was quite nice to see the stream. I could see stuff that normally I wouldn't see. *H31*

> I find it interesting because seeing it is the thing I have never done before because you can feel I know it is warm or something but I have never seen before it was very interesting and cool for me. *H91*

Additionally, participants reflected on how using the camera changed their perception of the surroundings. H21 and H61 remarked how using the camera introduced an adaption on how they see the world:

> Surprisingly even when I don't have the camera anymore I am using the hints I already got from the camera, it like how your eyes adapts to this kind of information. *(H61)*

> It makes us as human being sensitive to more than 3 colors. *H32*

## Potential for domestication

Participants reported that they would appreciate having a thermal camera in the home. They were comfortable with the mobile form:

> I was fascinated that you can have such an imaging tool as simple as a phone attachment. *(H13)*

However, they reported that the form factor depended on the use case. For instance, some preferred to have the thermal camera as a stationary tool monitoring a room. H22, wanted to have it as a substitute for the surveillance camera and he reported the need to have it in a stationary form:

> I can imagine if I have this camera in the ceiling I can monitor my sleep and my body temperature during the day to reflect my health.*(H13)*

One participant wondered if the technology will advance to have it in contact lenses for implicit vision extension. It's worth mentioning that participants also wanted to have it in different forms and communicated this idea with each other:

> It can be like Internet of things to communicate with anything. I want to see it in my cellphone, in my computer also. So it should be a separate thing but it has to communicate. *(H92)*

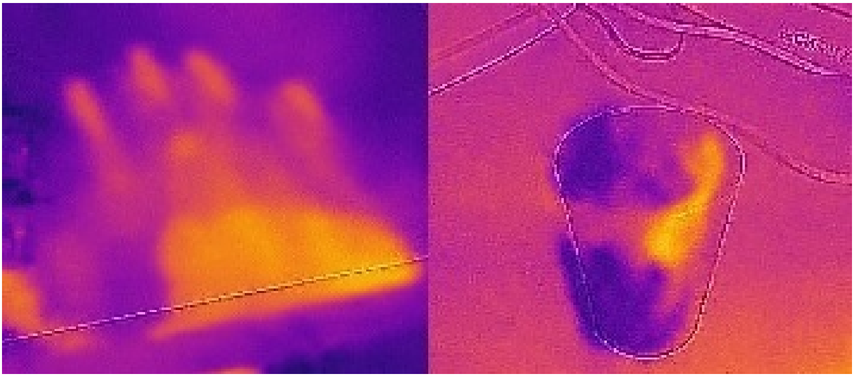


**Figure 3.6:** Traces when touching a surface and a cup.

## Social Implications

Most of the participants accept the idea of other people having thermal cameras operating at home:

> It is not problem if people want to use it. *(H7)*

Others reported being more careful to maintain the privacy of the house owner:

> I will be cautious. Maybe they will get upset if I touched their stuff as they can see when you touch something (Figure 3.6). *(H82)*

However, one participant reported to have privacy concerns:

> This could be a real privacy invasion so I would like to know if someone is having that advantage over me. *(H32)*

## When and Why do People use Thermal Cameras ?

Participants used the thermal camera in diverse tasks. Based on our analyses, the usage can be classified in four categories; *enhancing their awareness about the*

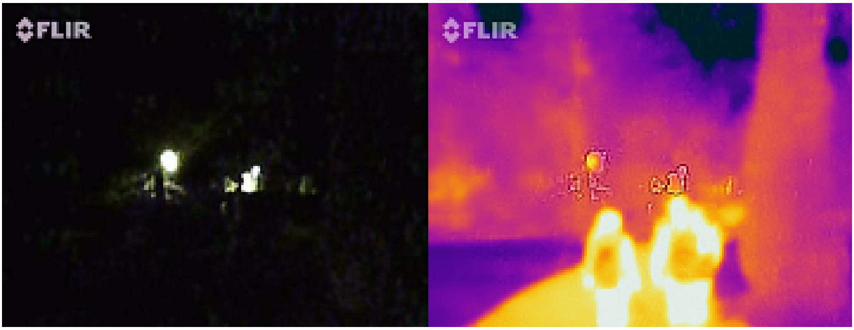



**Figure 3.7:** Thermal view (right), RGB view (left)

*environment, objects, people and substituting current techniques.*

***Environment Awareness***
Participants used thermal cameras to better understand and perceive their environment. For instance, H13 familiarized himself with the room's wall to know where to hang a frame away from the electric cables. H12 reported how he was informed about damage in his ceiling, at an early stage, saving him money:

> I figured out that my ceiling is damaged and fixing it at this stage saved me tons of money. Also I used to monitor if they fix it right and if the paint dried out. *(H12)*

The thermal cameras were also used in bad lighting conditions (Figure 3.7). For instance, H21 used it during the night in a dark room with her partner sleeping to find objects.

> I used the camera around the baby not to wake him up if I am looking for something I don't have to turn on the lights. And if I am looking for my phone during the night I don't have to turn on the lights and disturb everyone in the room. *(H21)*

***Object Awareness***
Participants also utilized the camera to know the state of different objects.

**Figure 3.8:** Sink trap blocked(left) and fixed (right).

> We had a problem with the sink so I took a picture, there was difference in colors. We called the specialist to fix it. At night I took another picture which had a clearer color and there was no difference in colors so we deduced that there was a problem and fixed now (Figure 3.8). *(H82)*

> We had an electricity cut out and I could tell from the camera that the freezer did not operate and it saved the food inside as well as the hassle of cleaning up after the food is destroyed, at that point I felt like we need this to be always monitoring the devices. *(H13)*

> I always open the dishwasher during its operation but I used it and found heat getting out from the dishwasher when its operating. *(H92)*

H11 remarked the ability of thermal camera to check her daughter's diaper.

> Checking if she needs to change I had to run after her and stuck trying to get her dressed again. It showed if the diaper has anything without even touching her. It saved time and money. *(H11)*

An interesting use case reported by one of the participants who was a refugee and had experienced war zones. He recognized the benefits of having a thermal camera in the home to detect hidden objects such as bombs. He also envisioned that the thermal camera could enable him to find monitoring devices such as hidden cameras and microphones as well as to find people. Additionally, he reflected on the ability of thermal cameras to detect invisible markers:

**Figure 3.9:** House from outside to identify opened windows or doors.

> They used to highlight the houses with wanted people using invisible paint, I can imagine I will be able to see it using this camera. I think having it might spare my life.

***People Awareness***
Participants also reflected the usage of the camera to know more about people's physical health state H92 or emotional state as reported by H11.

> I have a knee injury, it is always warm and if you hurt part of your body the injured part becomes warm (Figure 3.10) *(H92)*

> I know you can tell emotions from the temperature of the person, if I can hold the camera and know the unspoken feelings it kind of mind reading. *H11*

***Substituting Current Techniques***
Interestingly, participants highlighted the usage of thermal camera as an advanced substitution for existing techniques and technologies.

> I would like to use it as an early fire alarm, because the smoke detectors work when things are already on fire and smoke reaching the ceiling which is too late. *(H23)*


**Figure 3.10:** Knee injury (left), person's facial temperature (right)

> We have a surveillance camera with blind spots, with this camera I can see if someone passed by even if they are not around anymore (Figure 3.11). *(H12)*

> If you replace the Xbox camera with a thermal camera it would be a different experience. *(H22)*

H13 envisioned that a thermal camera could substitute health monitoring systems.

> If we can replace all vital sensing technologies with thermal cameras we could have a contactless health monitoring. *H13*

In summary, participants utilized thermal cameras in different use cases. Further examples include: kitchen tool, gaming camera, personal trainer, health monitoring tool and educational aid. It is worth mentioning that participants highlighted the need of applications to maximize the utility of thermal cameras.

> I think if you deploy it in form of apps like snap chat, fitness tracker, alarm systems its going to be really useful and maybe a necessity to have. *(H102)*

***Usage Behavior***
We analyzed the logged time of use. Participants used the thermal camera for an average of 20.4 min/day (SD: 4.7 min/day). A Pearson Product Moment correlation reveled that there was no relationship between usage duration and time ($r = -0.03$). This indicates continuous interest in using the technology.

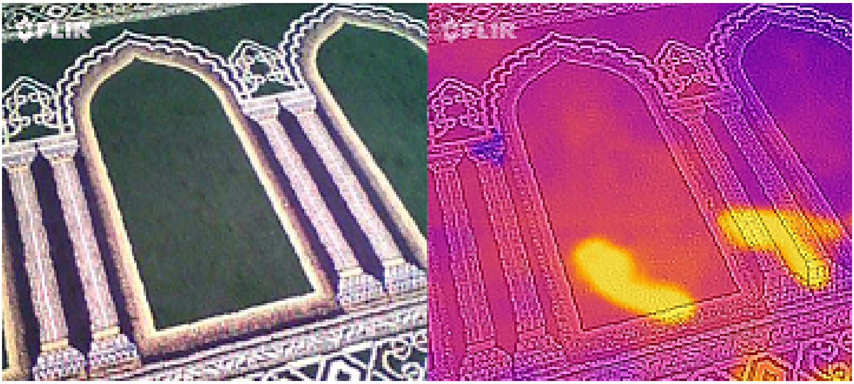



**Figure 3.11:** Traces of footsteps as surveillance technique.

| CATEGORY | EXAMPLE USE CASES |
|---|---|
| Enhanced Environmental Awareness | Checking ceiling for damages, checking walls for power lines before hanging a frame and finding objects in a dark room |
| Enhanced Object Awareness | Checking the state of the dishwasher, detecting issues with the sink, checking electric devices' states after electricity blackout, checking diapers and detecting invisible markers/paint |
| Enhanced People Awareness | Checking injuries and detecting emotions |
| Substituting Current Techniques | Early fire alarm, surveillance camera detecting past actions and replacing health monitoring devices |

**Table 3.3:** Summary of the usage categories.

### 3.2.3 Discussion

In the course of the conducted probe, we were able to gather insights from a total of 26 individuals who lived with thermal cameras for 10 days in their home settings. Here, we present insights drawn from our analysis of the results.

Our participants considered thermal cameras as a pleasurable tool and further considered it a vision extending tool that enabled them to see what they could not perceive with their own eyes. They used thermal imaging in the home for purposes including knowing more about their environment, determining the state of an object and emotions of people around them as well as to overcome the shortcomings of current technologies and techniques. They related their ability to see the temperature and thermal information to extending their visual perception as well as to compensate for environmental limitations. Finally, the extended ability made participants feel safe and having advantage over the environment and its limiting factors.

During our exploration, we encouraged the participants to envision use cases with no restrictions. As a result, participants proposed use cases applicable to high-end thermal cameras as well as an application layer, e.g. an application that detects the emotions. While many of the use cases are not yet applicable to the current state of the technology, this research suggest the need of future development of thermal imaging applications for everyday users. Our study enabled us to gather information from a diverse set of users. Furthermore, the home setups enabled the participants to envision the usage of extending visual perception in daily life. Participants reflected on how extended thermal view enhanced their perception and how it built new understanding of their everyday surroundings, where we observed that thermal imaging has a high potential for domestic appropriation.

Participants showed a clear understanding of the technology by exploration; there was no special training or instructions needed to understand and use thermal cameras. Additionally, they showed an awareness of the implications of the technology at hand, in highlighting the possible privacy issues that might arise during the usage of the thermal camera.

In summary, participants were excited about the use of thermal imaging to extend their visual perception at home. They envisioned potential use cases, as well as form factors. The collected insights reflect the importance of building and evaluating possible visual perception extension devices.

Concerning the future design of domestic thermal cameras, participants reported that the preferred form factor relates to the use cases, where a stationary form was preferred for always-on and environmental monitoring. However, they preferred the flexible portable form for real time explicit object exploration (e.g. checking diapers). Participants envisioned the wearable form factor as flexible and on-demand vision extension tool. These findings highlight that identifying ways to embed thermal imaging is an emerging challenge for HCI. Our exploratory study enabled us to gather information from diverse users; additionally the home setups enabled the participants to envision the usage of extend visual perception in daily setups. Participants reflected on how alternative vision modes provided a new information layer, enhancing their perception and understanding of their everyday surroundings. Interestingly, participants were eager to utilize the extended visual perception when their own vision was a limiting factor like in a dark basement with hidden objects and non-visible traces.

Participants used thermal imaging in the home for many purposes including enhancing their perception about their environment, the emotions and health state of people around them as well as to overcome the shortcomings of current techniques. In particular, participants reported that the thermal camera led them to detect sources of dissipating heat as hints for re-arranging objects in their house, spot dysfunctional domestic appliances, and locate and gauge sources of heating loss. These findings demonstrate the potential of the thermal imaging to reduce costs for household heating by detecting insufficient insulation and increasing energy saving awareness (e.g. checking if all windows are closed when leaving home). Surprisingly, in one case, the thermal camera was even reported as a tool that could potentially save/protect one's life in home settings when in a war zone. This shows that hardware miniaturization and democratization made it possible for thermal cameras (commonly only available to military personnel) to reach civilians, assisting them in escaping life threatening situations. Participants also reported the potential for eliciting emotions (e.g. anxiety) during interpersonal encounters, health state or disease symptoms (e.g. fever). Participants displayed increased awareness on the privacy implications of having such a layer of extra information at hand and commented on how it could be a potential means of discrimination (i.e. detecting and avoiding fevered peers). Although we anticipated participants would deploy thermal cameras in gaming as an alternative or addition to the Kinect depth technology, only one participant envisioned using it in a gaming scenario. However, it will be interesting to explore thermal imaging in the gaming context aiming to penetrate the gaming market in a similar way to how Kinect was introduced and is now adopted in a vast set of applications.

### 3.2.4 Summary

Through our review of related work, we concluded that no prior work explored the use of enhanced vision in everyday settings with the use of thermal imaging. Consequently, in this work we began our inquiry by exploring the users' perception of thermal spectrum and the potential use cases in daily life.

This research probe represents the first explorative study of the potential domestication and home uses of thermal imaging, in real homes by everyday users. While our findings are limited by the sample size and the technical limitations of the employed prototype, we were able to gather insights from a total of 26 individuals, who lived with thermal cameras for 10 days. Here, we present insights drawn from our findings.

Our findings imply the acceptance of the usage of thermal imaging. In the study we encouraged the participants to envision use cases with no restrictions, and they proposed use cases applicable to high end thermal cameras as well as an application layer, (e.g. an application that detects emotions). While many of the use cases are not yet applicable with the current state of the technology, this research suggests the need of future development of thermal imaging applications for everyday users and to identify implications for future extending visual perception systems using thermal cameras. These include identifying use opportunities, negotiating privacy, managing awareness and limiting dependence on enhanced vision.

We assured high ecological validity via the conducted technology-cultural probe in the home. We had the cameras provided in the home to allow creative and continuous ideation by the participants. In turn, participants highlighted their vision to have such a tool in their home to be used in a daily fashion. As a result we had insights concerning participants' preferences regarding the form factor. We hope our work can serve as an initial building block to understand what role enhanced vision can play in our future lives. We also hope that designers of future thermal imaging systems can use our insights to build enhanced vision interfaces with a high user benefit in everyday usage.

## 3.3 Chapter Summary

In summary, in this chapter we presented our exploration of the understanding of thermal imaging by two user groups; expert and novice. Our findings show that both expert and novice users of thermal imaging showed a high level of understanding of the presented thermal imaging technology: While the expert had a prior knowledge of the mapping, novice users acquired their understanding by exploring the scene and learning by doing (i.e. viewing). We concluded this chapter by presenting our reflection of the user reported experience with thermal imaging. In the next chapter we proceed to investigate the feasibility of the potentials of thermal imaging as well as the accompanied challenges.

q

# Chapter 4

# Amplified Perception of the Environment

Our eyes can only perceive light falling in the visible light spectrum, limiting the information perceived about our surroundings. In some cases, there are also hazards that tend to occur to us when we are not fully aware of our environment e.g. hot objects, gas leakage...etc. This is inevitable due to human's limited visual senses. However, as highlighted in the previous chapters, tools namely imaging technology paved the way for amplifying our perception, hence enhancing our awareness.

In this chapter we go along with the first investigation of deploying thermal imaging to amplify perception in both context-specific application (cf. Section 4.1) and daily setup (cf. Section 4.2). Through two study probes covering two different contextual scenarios, we explore and evaluate various facets of amplified perception. In section 4.1, we examine a firefighting context. We built two prototypes for FeuerWeRR. A hand-held devices, a sensing, a traditional way of utilizing thermal cameras by the firefighters. As well as, a HMDs alternative. Both prototypes provide the user with the thermal feed visualization.

The research questions we address in the presented chapter are:

- **RQ3**: How can thermal imaging be used to amplify perception of the environment ?

In Section 4.2 we present amplified environment perception in daily setup. Our environment holds a lot of cues that can inform us about our vicinity and enhance our awareness. Being aware of one's surroundings can enhance our awareness and the decision making process. By using *TriSight* a HMDs augmented with thermal camera and a depth sensor, we explore the feasibility of amplifying our visual perception using the current state-of-the-art technology. We provide abstracted feedback to the user in the form of the default color mapping of thermal feed. We ensured ecological validity by simulating daily setups.

Furthermore, section 4.3 investigates the thermal properties of different material that allows amplified interaction space, rather than the limited space by the field of view (FOV). Additionally, in this section, we investigate the influence of changes in the environment setup (indoors vs. outdoors) on the performance of the amplified interaction space.

In section 4.4 we explore the accompanied threats of the amplified perception. We investigate the feasibility of thermal imaging to infer users' smart-phone PINs and Patterns entry. Using a thermal camera and computer vision technique, we built *ThermalAnalyzer* to evaluate the performance of thermal attacks.

In the final section, We conclude with the a counter example of thermal attacks. In particular we introduce *VID* a vein based identification system, utilizing the ability of thermal cameras to capture the temperature difference between the blood in the veins and skin.

## 4.1 FeuerWeRR: Amplified Perception for Firefighters

*This section is based on the following publication:*

- Y. Abdelrahman, P. Knierim, P. W. Wozniak, N. Henze, and A. Schmidt. See through the fire: Evaluating the augmentation of visual perception of firefighters using depth and thermal cameras. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 693–696, New York, NY, USA, 2017. ACM

Firefighting is an activity that is highly dependent on the implicit knowledge and experience of the personnel involved. Firefighters build personal knowledge through training and service experience. The ordinary mode of operation of firefighters includes entering buildings full of smoke with limited visibility. Their task is to explore the environment and build a shared understanding of the situation, and decide on the next steps. Roles change frequently within the team to account for dynamic changes in the circumstances of the intervention. During an intervention, cognitive processing and time are key resources. Firefighters must actively manage their attention as they often perform multiple parallel tasks.

The availability of tools used e.g. thermal cameras provides information about the environment and offers opportunities for better coordination and improved safety. As a result the key resources time and cognitive processing could be supported. However, designing devices that can be appropriated and integrating them within existing procedures still remains a challenge. Designing devices for supporting time-critical collaborative work in hazardous environments is a recurring theme in HCI research [201, 202].

Mackay [161], identified design challenges for time-critical work. Kyng et al. explored the challenges of designing interactive systems for emergency response [145]. Previous work highlighted that designing for such complex environments confronts researchers with a set of constraints. Thus, understanding the ways to design devices and solutions for firefighting remains an open challenge [60].

As firefighter tools are mainly safety-critical systems, interaction techniques and methods that are suitable for this challenging environment are limited. Research conducted to date has aimed primarily at supporting commanders.

The work presented here is conducted within the scope of the FuerWeRR project. We focus on the human computer interaction aspects for designing for firefighters at the front line of an operation. The goal of the project is to build interactive tools for this complex environment that supports emergency operations, with a special focus on visualization modalities and environment perception. In our research, we consider the state of the art of thermal imaging technology and visualization being used by the firefighters. Currently, the firefighters use gray scale representation of the scene with minimal distance information. Hence, we aim to enhance and augment the presented information by designing and evaluating different visualizations and prototypes.

Our initial steps involved exploring how firefighters operate, current tools and technologies used and how they use and perceive these tools. We conducted a series of ethnographic methods including interview and participatory observation. In the following we discuss our findings and provide insights for future work.

### 4.1.1 Related Work

Firefighting is an activity that is highly dependant on the implicit knowledge and experience gained by training and real experiences. The common operation of the firefighters includes going into buildings full of smoke with limited visibility and their task is to explore the environment and build a shared understanding of the situation, as well as decide on the next steps. The role of the firefighters changes frequently within the team, and they have to deal with unforeseeable situations. Information resources are considered to be scarce in many ways, for instance the attention and processing time for available information are usually limited, as the firefighters must focus their attention on a set of tasks during operation.

The availability of tools used e.g. thermal cameras provides information about the environment; however, there is still the need of technology and tools appropriation process. While it might be considered to be straightforward to build devices to support firefighters, designing devices and solutions remains an open challenge [60]. There are limited interaction techniques and methods that are suitable for this challenging environment. As most of the research conducted aims to support commanders for instance, Landgren [148] investigated evolutionary design of large displays for collaborative work between commanders [125].

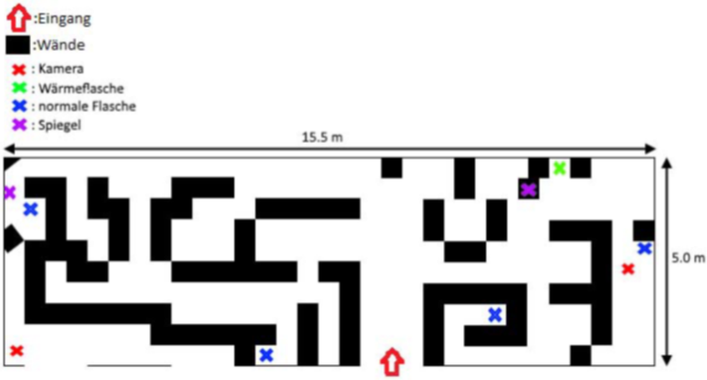



**Figure 4.1:** Map of the Basement.

Researchers identified communication patterns, highlighting the importance of verbal communication [149], and suggest system to enhance and improve communication practices [43]. On the other hand, work involving the firefighters in action usually focuses on the technology rather than the interaction aspect [60]. Recent work [176] focused on the interaction aspect of firefighting, however it does not consider all the collaborative aspects of firefighters' operations.

There is research conducted relevant to this work –Designing devices for supporting time-critical collaborative work in hazards environment. Mackay [161], identified design challenges for time-critical work. Other researchers further explored the challenges of designing interactive systems for emergency response [145]. Designing for such complex environments confronts researchers with a number of constraints. In this section, we present our findings of the studies that we conducted with the firefighters, including lab studies, interviews, and a survey.

### 4.1.2 Study I: Depth Vs. Thermal Vision Amplificat

In our first study, we aimed to assess the usage of the amplification mode; in other words the benefit of augmenting depth as opposed to thermal information. We used a simulated challenging fire environment, by using a dark basement with obstacles as depicted in figure 4.1.

### Participants and Procedure

We invited 16 participants aged between 20-30 years; 12 males and 4 females. None of the participants had been to the basement before or had prior experience with depth cameras.

We used the hand held prototype shown in figure 4.2, where participants could change the views either using the depth or thermal view. We focused on a navigation task, as this is one of the most challenging tasks of firefighters. Participants were asked to perform two tasks:

- Estimate the size of the basement
- Find the warm bottle

After task completion, the participants were interviewed to rate the usage of the depth and thermal camera to perform the tasks. Additionally, we logged the view they used to perform the tasks as well as the camera feed.

## 4.1.3 Results and Discussion

We analyzed the views and the recorded feed from both cameras to identify the camera used to perform the tasks. We found out that 87% of the participants used the depth camera to perform the navigation task to estimate the size of the basement. On the other hand, all participants switched to the thermal view to locate the warm bottle.

Our findings from the interview confirms our findings. All participants reported that they had an enhanced experience using the extended views and the difficulty of the tasks was drastically reduced. Additionally, six participants recommended the usage of both imaging technologies in the same view through combining the depth and thermal information via sensor fusion.

The first study reflects the importance of both imaging technologies, where extending the visual perception to the depth band allowed better navigation skills, and the thermal layer allowed sensing non-visible information (i.e. temperature of the bottle) as well as performing fine search tasks. Accordingly, we performed a second user study with three different views: (1) Thermal, (2) Depth and (3) Thermal and Depth fused view. Additionally, we used two different form factors.

**Figure 4.2:** HMD (left) and Hand-held (right) prototype used by firefighters.

### 4.1.4 Study II: Hand Held Vs. Head Mounted Display

The firefighters are used to the hand held form given it is the typical form of the currently used devices. However, we wanted to evaluate different form factors, namely hand held and head mounted Display.

We invited 11 participants aged between 19-51 years. None of the participants had been to the basement before and all had no prior experience with depth cameras. Six participants used the HMD and the other five used the hand held setup. The participants were asked to perform the same tasks as in study I. We logged the time spent in each view of the thermal, depth and fused views. Additionally, we conducted post-interviews with the participants.

### 4.1.5 Results and Discussion

We computed the time spent in each view, and found that 39% of the time was spent using the thermal and depth fused view, followed by 35% and 26% in the depth and thermal views respectively. This confirms the effectiveness of the fused view, and the usage of both imaging technologies to maximize the augmentation of human visual perception. Additionally, the interviews reflected the preferences for using the HMD. However, all participants mentioned that the prototype should be integrated in the firefighters' helmets to be usable in real fire scenarios.

### 4.1.6 Summary

We presented two studies conducted with firefighters to investigate the potential of augmenting and amplifying their visual perception using both depth and thermal cameras. We covered in this section the evaluation of the potential form factors for amplified vision, using HMDs and hand-held. We aimed to evaluate the impact of using thermal vs. depth; and additionally, the preferred form factor. Our findings reflect the effectiveness as well as the acceptance of the amplifying prototypes, where all participants reported an enhanced experience while using the prototypes. Moreover, participants from study I recommended the use of the combined thermal and depth fused view. Hence, we modified them to include the fused view in study II, and it was the most used view confirming the reported user preferences. We also found out that the firefighters preferred the HMD to the Hand-held, due to its hand-free operation.

## 4.2 TriSight: Amplified Perception for Novice Users

*This section is based on the following publication:*

- Y. Abdelrahman, P. Wozniak, P. Knierim, N. Henze, and A. Schmidt. Exploration of alternative vision modes using depth and thermal cameras. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, pages 245–252, New York, NY, USA, 2018. ACM

As shown in the previous section, imaging technology has shown promising advancement in visualization and perception of the environment, for specific contexts such as firefighting. Recently, thermal imaging (far-infrared) has become affordable for commercial and personal use in terms of cost and size, to the extent that thermal imaging has been integrated into some mobile phones, enabling an extension of how we perceive, visualize and interact with the surroundings. Thermal imaging provides a heat map of a scene to the user in a contact-less, real time, and robust manner (e.g. it is light independent). Moreover, infrared imaging (including both near- and far-Infrared) enhances and extends the perception of our visual sensing. By using these cameras, we are capable of perceiving light outside the limited visible spectrum. Concurrently, HMDs and augmented reality (AR) have shown significant advancement in visualization and interaction challenges in multiple domains, e.g. gaming, medicine, and manufacturing. However, despite the advancement of AR applications, it is under-explored in daily life.

In contrast to traditional RGB cameras integrated in HMD and glasses, the design and integration of infrared cameras, especially in daily usage, have not been explored yet. We investigate the potential of thermal and depth imaging for daily use. We build an initial prototype to explore the feasibility of vision extension and a switching system with currently available technology. We conducted a study using the prototype to explore task scenarios and how users switch between vision modes. We determined practical and technical challenges for the development of tools that augment visual perception. We are using a *TriSight* proof-of-concept prototype and created situations and environments. This work aim to inform technology and human factor design decisions towards a real interactive vision extension system for evaluation. We chose our approach to achieve the fullest involvement of the users in the design of the vision amplification systems.

### 4.2.1 Related Work

Previous research proposed several systems that enhance human visual perception in special scenarios and setups. In the following, we present prior work in research aimed to provide an extended visual perception using depth and thermal cameras.

#### Extending visual Perception

Research has been done to extend our visual perception, for instance Kimber et al. [135] used mirrors to augment user perception. Others looked into extending the perception via extending the field of view [78, 180, 181]. Recent research explored the visualization of electromagnetic radiation emitted from wireless technology such as Wi-Fi hotspots or cellular towers [79, 95].

One of the limiting factors in visual perception is the ability to infer exact depth information, for instance for navigation and obstacle avoidance. Biswas and Veloso used depth information for obstacle avoidance for autonomous navigation [36]. Their system works on the depth information alone and does not require RGB data. Izadi et al. presented KinectFusion, a system that allows users to hold a Kinect to generate 3D models of a scene in real-time [121]. Researchers argue that 3D maps have many applications, ranging from elderly support [286] robotics, telepresence, gaming and AR. They further state that they believe that the price of RGB-D cameras will decrease in the future, allowing the cameras to enhance different kinds of application scenarios, as well as human perception. Depth cameras have been explored and utilized in various applications and domains. However, there is limited research investigating the domestication of depth cameras beyond gaming.

Concurrently, thermal cameras operating in the far-infrared spectrum and can be used to extend the spectrum that is perceivable. Researchers highlight that thermal cameras are continuously getting cheaper and more and more cameras are appearing on the market. This enables thermal cameras to be used in a diverse set of applications, by enhancing existing application scenarios and opening new ones. Matsumoto et al. use KinectFusion and combine RGB-D camera with a thermal camera. This system allows a visualizing the thermal distribution in the environment [159, 168]. Vidas et al. combined a range sensor with a thermal-infrared camera in a hand-held system to generate dense 3D models of building interiors. Combining a low-cost RGB-D camera and a thermal-infrared camera enables generating 3D models that contain surface temperature information [247]. Van Baar et al. highlight how thermal cameras can be used to improve object segmentation [243]. Using thermal cameras has also been explored for expert users. Previous work explored how to extended visual perception in special situations and environments. For instance, it has been targeted for firefighters usage [6, 21, 22]. On the other hand, there is limited work on the usage of imaging technologies for novice users.

Past research shows that while users can benefit from additional vision modes, solutions that effectively deliver additional vision information are yet to be delivered. Thus, understanding if and how interactive systems can convey content beyond the regular vision spectrum emerges as a challenge for HCI. Here, we perform an initial exploration of this design space by building a proof-of-concept prototype and evaluating in a study. We conducted a mixed-methods lab study to explore initial reactions to an extended vision system and investigate its potential for domestication. Consequently, we explore the What is the potential for domesticating augmented vision systems as well as What are user attitudes and design challenges for developing systems that support alternative vision modes.

### 4.2.2 Trisight: A Proof-of-Concept Prototype

We built *TriSight* that extends the user's visual perception. It enables users to perceive the environment in three different modalities; visual view recorded through RGB cameras, depth view recorded by a depth image sensor, and a heat view recorded by a thermal camera. Through the attached different cameras; RGB, Depth and Thermal cameras, the HMD presentes the camera feed to the user, where the RGB is used to allow having a see-through experience, to be able to see the environment, the depth and the thermal viewed the heat information of the environment. Users can select one view at a time using a wireless controller depicted in Figure 4.3. Users can freely switch back and forth between the three views with a simple click. *TriSight* puts three different image sensors into service. We use two Matrix Vision mvBlueFOX-MLC202b cameras to enable stereo vision. The cameras operate at a resolution of $1280\times960$ pixels with a FOV of $118^\circ\times87^\circ$, running at 90 frames per second (FPS). For thermal imaging, we used an Optris PI450 camera with a resolution of $382\times288$ pixels. The spectral range is between 7.5 and $13\mu$m with a noise equivalent temperature difference of 40 mK at 80 FPS. The thermal camera is equipped with a 7.3 mm lens with $80^\circ$ horizontal and $58^\circ$ vertical FOV. The third sensor is the structure sensor from Occipital [15]. It operates with a resolution of $640\times480$ pixels and FOV of $58^\circ\times45^\circ$ at 60 FPS.

[15] www.structure.io/

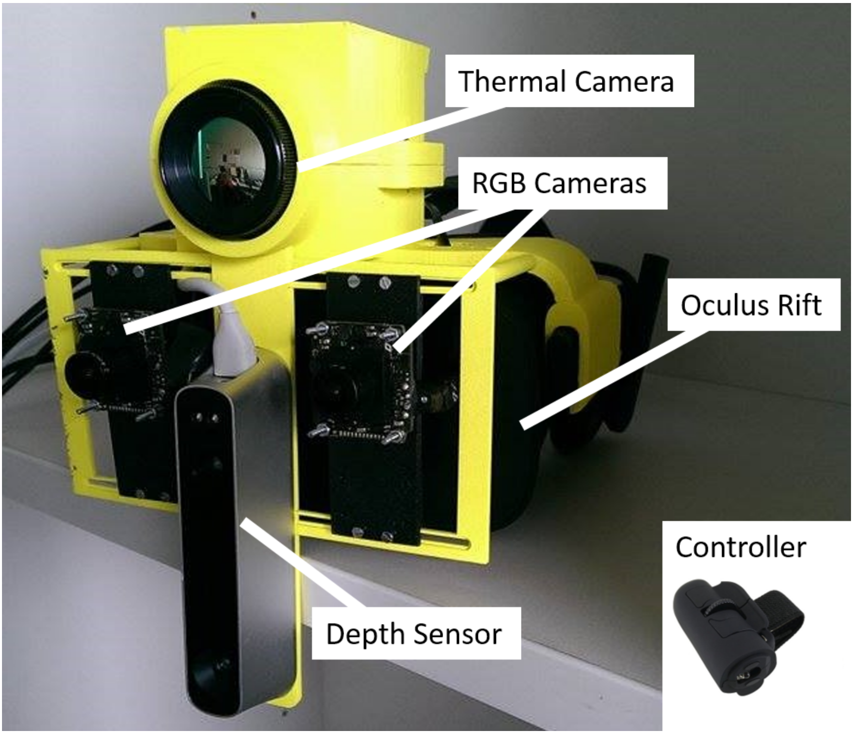



**Figure 4.3:** TriSight: OculusRift, two RGB cameras and a thermal camera.

All image sensors are attached to a custom designed printed chassis (yellow part in Figure 4.3). This frame connects all sensors with the Oculus Rift CV1 which we used as the output device. We ordered the sensors in a way that enabled the largest view overlap as shown in Figure 4.3. The dimensions of *TriSight* were 20×16×14 cm (w×h×d) and total weight of 1.2 kg including the HMD. Both the Oculus Rift and all imaging sensors were connected to a MSI-GT72 [16] notebook. The notebook powered the attached hardware and did all the processing. The notebook was placed in a backpack.

[16] https://www.msi.com/Laptop/GT72-6QD-Dominator-G.html

We used the Unity 3D game engine for image stream processing since this offers powerful Oculus Rift support. The image processing pipeline consists of three steps. Each stream from one of the camera systems is loaded separately, cropped to compensate the different FOVs and lastly rendering a split-screen stereo with distortion correction for each eye. Only one stream is visible at a time. The user can switch between the streams with a press of one of the 2-buttons of an ergonomic finger mouse (controller in Figure 4.3).

### 4.2.3 Evaluation

We used a proof-of-concept prototype in created environments aiming to achieve the fullest involvement of the user's in the design of the vision extension systems. It allows the investigation and evaluations of user experience, and makes qualitative as well as quantitative measurements of usage, e.g. the time spent in each view.

The study was comprised of two parts. In the first part, we studied the usage of the three different views in daily tasks, by creating three lab environments: kitchen/home, office and basement. In the second part, we interviewed the participants for usage insights.

1. Office Tasks:
   (a) organizing and archiving files.
   (b) identifying connected plugs.
2. Basement Tasks:
   (a) locating an object
   (b) detecting leakage
   (c) detecting a disconnected pipe
3. Kitchen & Home Tasks:
   (a) finding a pet
   (b) checking a plant
   (c) preparing a hot drink

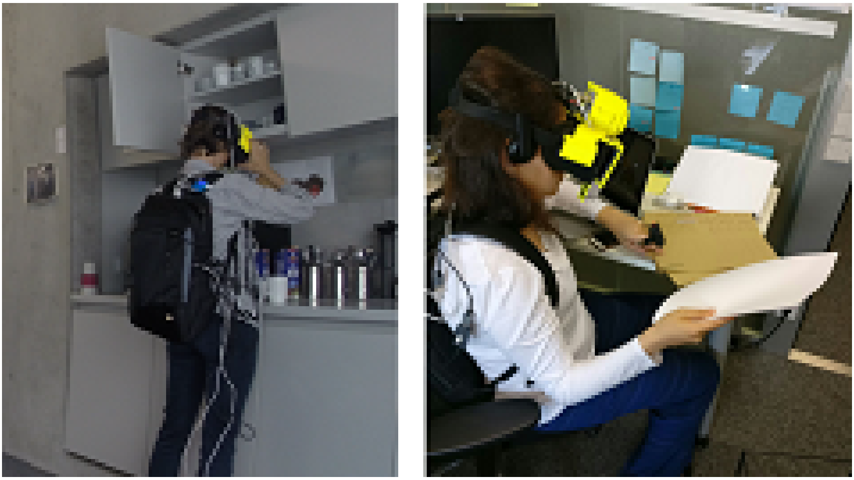


**Figure 4.4:** Participants while performing kitchen and office tasks.

The second part was a post-task interview evaluating their experience with *TriSight*. We conducted the interview in the lab after the participant explored the three home-like setups. In the interview, participants were asked about their experience with *TriSight*, and how a similar device could be used in their daily life. Additionally, they reported their reasoning for the view selection to perform the task. The post-task interviews were recorded for later transcription and thematic analysis, to better understand the content of the interview.

## Participants and Procedure

We recruited six participants (3 female, with an average age of 29.17 years, SD=10.83). Participants had a diverse background; we had one secretary, one doctor and four students in different majors. Two participants had previous experience with augmented reality, one had experience with depth cameras and none had any experience with thermal cameras.

After welcoming the participants in the lab, we asked them to sign a consent form and explained the purpose of the study. The experimenter then asked them to fill in a questionnaire concerning their experience in using HMDs, RGB and thermal cameras. Throughout the study we recorded the views displayed to the participants as shown in Figure 4.5.

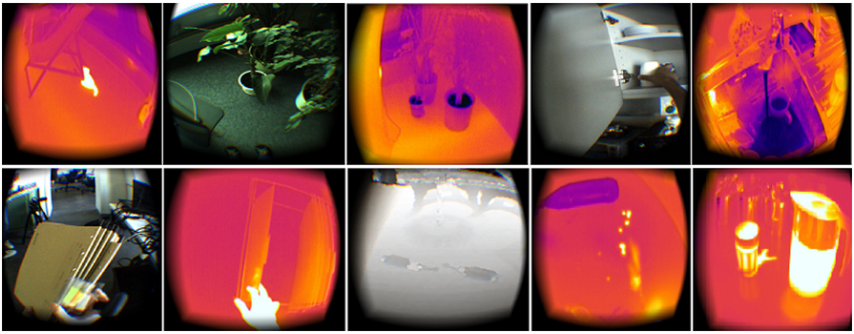


**Figure 4.5:** Images taken with *TriSight* by participants during the study.

We recorded their state using the Positive and Negative Affect Schedule questionnaire(PANAS) [262] before the start of the study and after each environment. Moreover, we logged the time spent in each view. The participants were encouraged to speak their thoughts out loud, and comment on the views used to perform the tasks. These notes where recorded by the experimenter for later analysis. The order of the environments was counterbalanced using a Latin-square. The study took approximately 120 minutes ( 30 minutes in each environment and 30 minutes for the post-task interview) per participant. The whole experiment was recorded using a Gopro Hero3 RGB video camera.

### 4.2.4 Results

We collected 540 minutes of video: 199.53, 158.51, and 181.56 minutes in RGB, Depth and Thermal views respectively. Further, 163 minutes of qualitative interviews were collected along with the comments recorded during the study from the participants. We analyzed the PANAS scores after using *TriSight* in the different environments, (2) time spent in each view, (3) the views used to perform the given tasks and (4) the conducted interviews.

### PANAS

We assessed the positive and negative affect at the beginning of the experiment as an indicator for the initial/baseline of the user's state. Scores were recorded

| | Positive Affect | Negative Affect |
|---|---|---|
| Baseline | 23.83±5.93 | 21.67±4.07 |
| Office | 26.00±8.29 | 21.17±5.67 |
| Basement | 29.33±9.57 | 21.00±6.44 |
| Home & Kitchen | 29.00±7.96 | 21.17±5.54 |

**Table 4.1:** Mean and standard deviation of the PANAS scores.

after each environment aiming to assess the experience of *Trisight* and whether it influenced the user's state negatively or positively. As shown in Table 4.1, participants' overall positive affect increased and their negative affect score dropped.

## Time Spent in Each View

We used the logging of switching between views to compute the time spent in each view. Figure 4.6 presents the percentage of time spent in each view. In the home and office environments, participants spent most of the time using RGB view 42% and 47% respectively. However they still spent time viewing the surroundings with thermal and depth view.

However, when the information provided by their visual sense failed to provide sufficient information they tended to switch to the other views as depicted by the time spent in the other views. For instance, participants used the thermal view to check the soil of the plants to know if it had been watered or not, others used it to search for the pet, by utilizing the heat emitted from the cat-shaped bottle.

Interestingly, participants utilized the depth and thermal views in the situation where the normal RGB view has limited capability. As shown in table4.1, in the basement, where there was limited light, they only spent 23% of the time in the RGB, and the rest was spent in the depth and thermal views, 34% and 43% respectively. As observed in the views used to perform the tasks and time spent in each view, thermal imaging showed a higher potential as an alternative vision mode.

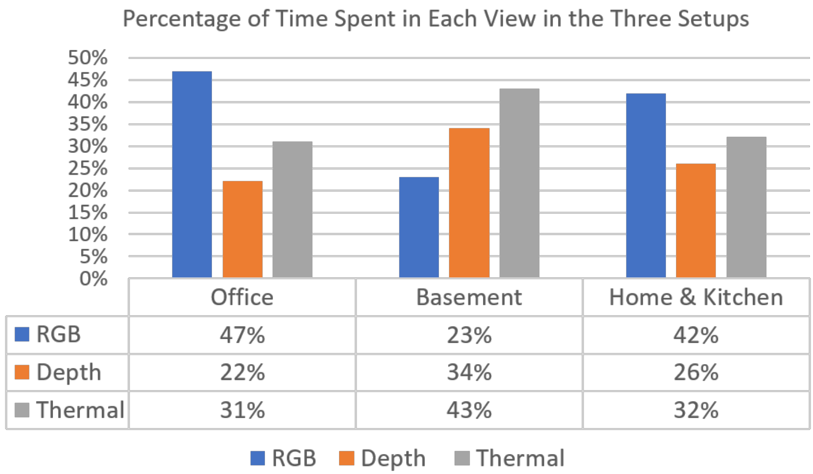



**Figure 4.6:** Time spent in each view, where the total time is 30 minutes.

## Views used

Participants were instructed to name the views used to perform the tasks in each environment. The experimenter recorded the views used to perform the tasks, aiming to access the potential usage of the presented views. Figure 4.5 depicts examples of the different views used to achieve the tasks. As shown in table 4.2, participants used different and sometimes combined views.

Confirmation for the time spent in each views, the RGB was mostly used in the familiar environment (office and home), however, participants also utilized other views. For instance, one participant used the thermal view to identify the last used file, utilizing the heat trace left behind. Anther used it for detecting if the plants need to be watered based on the temperature of the soil. One participant used the thermal and RGB views to prepare a hot drink, viewing the hot spots of the cup and the water to better handle it. The comments of the participants as well as the views used reflect the potential benefits of extending the human visual perception beyond the limited light spectrum.

| | Home & Kitchen | | | Office | | Basement | | |
|---|---|---|---|---|---|---|---|---|
| | Find the Pet | Check Plant | Prepare Drink | Organize Files | Identify Connected Plugs | Locate an Object | Leakage Detection | Find Disconnected Pipes |
| P1 | Thermal | Thermal | Thermal+RGB | RGB | Thermal | Thermal | Thermal | Depth |
| P2 | RGB+thermal | Thermal | Thermal+RGB | RGB | RGB | Depth+Thermal | Thermal | Thermal+Depth |
| P3 | Depth+RGB | RGB+Thermal | RGB | Thermal+RGB | RGB+Depth | Thermal | Depth+Thermal | Depth |
| P4 | Depth+thermal | Thermal | RGB+Depth | RGB | Thermal+RGB | Thermal | Thermal | Thermal |
| P5 | RGB | RGB+Thermal | RGB | RGB+Thermal | RGB | Depth | Depth+Thermal | Depth+thermal |
| P6 | RGB | Thermal | Thermal+RGB | RGB+Thermal | RGB | Depth+Thermal | Thermal | Thermal |

**Table 4.2:** Views used by participants to perform the given task.

## Interviews

The post-task interviews were recorded and transcribed for analysis. We used thematic analysis to better understand the content of the interview. Overall, 163 minutes of recordings were analyzed. Two coders coded 15% of the corpus independently using nVivo [17]. Afterwards, they met to assess differences and constructed the final coding tree. The rest of the corpus was coded by a single coder. The final themes emerged from the coded quotations in a final session with two researchers. We present the four themes below.

## Future form factor

The users noted that our current prototype was rather bulky and difficult to wear over extended periods of time. However, participants were also eager to envision what a future device offering multiple vision modes could look like, and what they would see as key features of such a device. P3 reflected that effectively using multiple vision modes would require highly developed wearable technology:

> [It should] look like normal glasses, light weight, the setup doesn't need much space or doesn't have problems with mobility. It doesn't drain power. *(P3)*

Participants also anticipated that devices that offer alternative vision modes will soon be offered as smartphone accessories. One participant assumed this and speculated on how the device would connect with their mobile device

> It has to be smaller, lighter and something wireless. The data can be send over... wirelessly over Bluetooth or the Internet for example. *(P1)*

[17] `http://www.qsrinternational.com/nvivo-product`

## Reflecting on future use

Participants were eager to speculate on how alternative vision modes could be useful in their future everyday lives. P2 commented they often struggled with leaks and excessive humidity and would benefit from easily accessible thermal vision. P5 noted that they would happily turn on thermal vision when under stress, in order to identify objects quickly:

> It would be useful to have a device like that in stress situations where I have to find something very quickly, when I am under time pressure. *(P5)*

Participants remarked that tangible benefits of introducing alternative vision modes in specific settings could result in swift acceptance. They often related the new capabilities of our prototypes to tasks they perform regularly. For instance, P6, a physician, remarked:

> I would use it in the future, as it is a good additional way to examine my patients. *(P6)*

## Sensing the environment

Users often reflected on how they understood the different properties of the environment that they could perceive by using multiple vision modes. They were eager to illustrate possible benefits with scenarios:

> If I were on holiday I could check the water temperature and decide if I want to swim. I could look at a bench that's far away and check its temperature, if it were super hot I wouldn't go over there. *(P2)*

Similarly, they appreciated how alternative vision modes allowed for enhanced perception of environments with which they were already familiar. The participants wondered how additional vision modes could build their new understanding of their everyday environments:

> [in the basement] I could finally know how these pipes are connected. *(P4)*

### Social aspects

Finally, users wondered how head-worn enhanced vision devices would affect social encounters. One participant recognized that thermal images may be linked to one's physiological state and thus may generate privacy issues:

> Maybe it will generate another [type of] discrimination — I walk down the street and I have a fever so people will try to avoid me. *(P2)*

Some participants were also worried that thermal vision may create issues related to gender and sexuality

> It will embarrassing, because, I think, if wear something thin, people will be able to see through. *(P2)*

Furthermore, participants agreed that others should not be made aware if one were using augmented vision. One participant remarked that transparency was not an issue:

> In the future, maybe, everyone will be using it on daily basis, so we shouldn't show [what mode one's using]. *(P1)*

While another believed that showing the state of an augmented vision device was needed as it is only a sensor:

> I don't think they need [a way to show status] as it has nothing to do with other people, it won't invade others privacy or harm them. *(P3)*

## 4.2.5 Discussion and Summary

Our study demonstrated that augmented reality technologies may be appropriate for communication augmented vision information. We observed that non-expert users were able to easily understand and utilize the way thermal cameras work through an HMD. As users reported no major difficulties performing the tasks and exploring the environment, we can conclude that AR offers the potential for easy and fast deployment of thermal imaging. Further, HMDs may reduce the need for training. Further, we found that *TriSight* improved the participants' subjectively reported affective state as reflected in the PANAS scores. This indicates that an amplified vision technology may be perceived as beneficial while performing domestic tasks, thus showing potential in amplified perception.

Our study also revealed a number of opportunities and challenges for future development of augmented vision systems. As we observed in the views used to perform the tasks and time spent in each view, thermal imaging showed a higher potential as an alternative vision mode. Thus, future designs should explore how to offer an experience balanced between different vision modes and offer vision mode changes when appropriate. The usage of different views reflects the participants' understanding of the other spectrum bands, although most of them did not have previous experience with this technology. Interestingly, all participants used the extended vision in the basement. It appears that they used extended vision when the RGB (i.e. our visual perception) shows limited capabilities as reflected in the time spent in each view, as participants spent more time in using the depth and thermal cameras in contrast to the RGB. Thus, we see an emerging need for finding ways to communicate the properties of different augmented vision modes to users effectively to foster their implicit awareness of enhanced perception available. An intuitive understanding of alternative vision modes appears to be a necessary condition for effortless vision mode switching.

Where we only considered explicit vision mode switching, our findings highlight the need to explore implicit and context-aware switching techniques, for instance based on the lighting conditions. As the explicit switching relies on the understanding of the user of the technology in hand, having an implicit mechanism would enhance the understanding of the imaging technology as well as assist in the appropriation of the extended vision.

In summary, our findings highlight the potential of alternative vision modes. Participants reflected on how these modes provided a new information layer, enhancing their perception and understanding of their everyday surroundings. Interestingly, they were eager to utilize the extended visual perception when their own was a limiting factor like in a dark basement, with hidden objects and non-visible traces. However, researchers and designers must consider the challenges and considerations while designing such vision amplification tools.

## 4.3 ThermalMirror: Amplified Interaction Space

*This section is based on the following publication:*

- Y. Abdelrahman, A. Sahami Shirazi, N. Henze, and A. Schmidt. Investigation of material properties for thermal imaging-based interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 15–18, New York, NY, USA, 2015. ACM

In the previous sections, we presented the work done in amplifying our visual perception. Another modality to consider is the non-visualized amplified environment, where the thermal view is not presented to the user but rather to the system to allow amplified interaction space. As a novel sensing technology for human-computer interaction, thermal imaging offers exciting opportunities for the development of interactive systems. Prior work investigated the usage of thermal imaging to amplify the interaction space as depicted in Figure 4.7 [210]. Building upon our previous work, in this section we investigate the material properties that enables such amplified interaction space.

In this section, we present a study that assesses the recognition accuracy of mid-air gestures sensed through thermal reflection using different surfaces in indoor and outdoor setups to complement the work presented by Sahami Shirazi et al. [210]. Further, we provide holistic insights on the surfaces' properties that should be considered when it comes to sense interactions on the surface and/or mid-air using a thermal camera. Based on the identified properties, we provide a material space describing surfaces that enable interaction on the surface and/or mid-air interaction through thermal reflectivity. The guidelines enable to select materials for creating interactive surfaces using thermal imaging.

### 4.3.1 Related Work

Enabling natural gestural and touch-based interaction techniques is one of the main goals of human-computer interaction research. In the following, we discuss prior work in three different areas aimed to provide natural interaction using (1) thermal imaging, and techniques used for recognizing human gestures to interact with (2) stationary as well as (3) mobile projected screens.

Thermal imaging has recently also been used as a sensor for enabling the interaction with arbitrary surfaces [118, 119, 150]. This is achieved by integrating thermal imaging and existing computer vision techniques to improve user surface interaction by utilizing advantages of thermal imaging to overcome common RGB and depth cameras' drawbacks. Daisuke and Koskue, for example, used short-lived heat traces that result from the heat transfer from one object to another as means for interaction and created an interactive tabletop surface [118]. Larson et al. used heat traces caused by fingers touching a surface and detected pressure for interaction on surfaces [150]. They reported that the thermal reflectivity of surfaces induces noise for their system.

Touch and mid-air gestures are common techniques to interact with projections. These are typically detected using either RGB, infrared (IR), or depth cameras. There exists a large body of work focusing on detecting and tracking hands and fingers to enable multi-touch and mid-air gestures using RGB cameras [48, 137, 127, 166]. Such systems typically use skin color detectors [127] or template matching [137] to segment the hand and then calculate contour and convexity defects [166] to identify fingers.

Infrared imaging is a popular technique to enable multi-touch and mid-air gesture when interacting with projection screens [107, 105, 120, 267, 268]. In such systems, the space behind the screen is typically illuminated with an infrared source and all except the infrared light is blocked from the camera using an infrared-pass filter. This technique has been widely used for tabletop interaction by combining a rear-mounted IR camera and a projection unit. Using the depth map provided by depth cameras is another approach for detecting touch and hand gestures on projected screens [175, 269, 270, 271]. These systems generally utilize either a 2D view above the surface [175, 270] or a selective 2D projection of 3D sensed data [271] for processing users input on or above the surface using common 2D computer vision techniques.

Our research is also related to previous work on interaction with mobile projectors. According to the work by Rukzio et al. [209], interaction with mobile projected displays can be divided into four categories. A common approach is to separate input and output and use the touchscreen of a mobile phone [94] or a touch sensor for input [38]. Researchers have also investigated input for mobile projection by moving and gesturing with the projector itself [38, 45, 266] or by aiming with the projector at objects in the environment [218]. Although these solutions allow users to focus on the projection and perform intuitive gestures, tracking the projector's movements requires additional hardware equipment to be installed in the room or mounted on the projector unit.

Directly touching the projection screen with the fingers [273, 100] or using a stylus [44] is another approach to interact with mobile projections. However, such a setup requires users to be very close to the projection leading to a small projection area and large shadows on the projection. Another solution is to use mid-air finger pointing and hand gestures to interact with the projection [172, 51]. The SixthSense system [172], for example, offers a set of mid-air hand gestures to support interaction with the projection. The system uses a color-based approach to track fingers. ShadowPuppets [51] provides shadow gestures as input to a handheld projector system by casting hand shadows for co-located collaborative scenarios. Winkler et al. [272] have argued that it is preferable to perform gestures besides or even behind the projector. This is not possible with current handheld projector-camera systems as both face the same direction and the projection and sensing spaces overlap. Thus, users occlude the projection while performing mid-air gestures in front of the camera. A thermal camera, however, can detect direct interactions on a surface and mid-air gestures as well as a user's interaction out of the camera's direct field-of-view, for instance behind the camera.

A large body of work combined projectors and sensing cameras to build interactive projected surfaces. Initially, RGB cameras were used to detect hands and fingers [127, 166]. Such systems typically use skin color detectors or template matching to segment the hand and then calculate contour and convexity defects to identify fingers. A major challenge of such systems is the sensitivity for different light conditions. Research prototypes have used infrared imaging [107] and depth cameras to enable multi-touch and mid-air gestures when interacting with projection screens and tabletop setups. In such systems, the space behind the screen is typically illuminated with an infrared source. Using an infrared-pass filter, all lights except the infrared light are blocked for the infrared camera. Using a depth camera, the depth information can be used to detect touch and hand gestures on projected screens [175, 271]. Such systems generally utilize either a 2D view above the surface [175] or a selective 2D projection of 3D sensed data [271] for processing users' input on or above the surface using common 2D computer vision techniques.

Using existing computer vision techniques, thermal imaging can be used to detect interaction on surfaces as well as mid-air gestures. Two thermal properties have been leveraged for monitoring interactions. First, heat traces that are left on a surface due to the temperature difference and the heat transfer between hands and surfaces. Such traces have been used to detect interactions and pressures

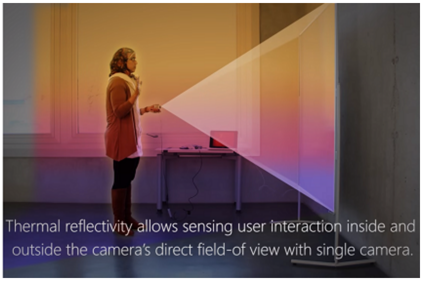




**Figure 4.7:** Amplified Interaction Space

on surfaces [119, 118, 150].The second property is thermal reflectivity which is the result of radiation reflection when striking a surface. Sahami Shirazi et al. [210] propose to use specular reflectivity for extending the mid-air interaction space behind the camera's direct field-of-view and detecting mid-air gestures. However, they did not report the recognition accuracy of mid-air gestures in the extended space. In contrast, we investigate the recognition accuracy of mid-air interaction using thermal reflectivity. Further, we provide a holistic overview on surface properties which should be considered for creating an interactive setup using thermal imaging as the sensing technology.

## 4.3.2 Thermal Reflectivity

Thermal radiation, as a result of energy transitions of molecules, atoms, and electrons of a substance, is continuously emitted by all matter whose temperature is above absolute zero. The spectrum and intensity of black body radiation depends on the object's temperature as expressed by the Planck's and Stefan-Boltzmann laws. The radiation emitted by objects at room temperature falls into the infrared region of the spectrum, which extends from 0.76 to 100 micron. The human body's net radiation is, for example, around 142 watts (W), with a skin temperature of 33°centigrade (C), at an ambient temperature of 22°C, and a peak wavelength of 9.5 micrometer ($\mu$m).

When radiation strikes a surface it is partially absorbed, partially reflected, and the remaining part, if any, is transmitted. Based on the first law of thermodynamics the sum of absorbed, reflected, and transmitted radiation is equal to the incident radiation. For fully opaque surfaces the transmissivity is zero, thus, the sum of absorptivity and reflectivity is one. The absorptivity is independent of a surface's temperature. However, it depends on the temperature of the source at which the incident radiation is generated. The reflectivity of a surface depends not only on the direction of the incident radiation but also on the direction of the reflection. Surfaces are assumed to reflect in two manners: specular and diffuse. In specular (mirror-like) reflection, the angle of reflection equals the angle of the radiation beam. For diffuse reflection the radiation is reflected equally in all directions regardless of the incident radiation's direction. The reflectance of a surface depends on its roughness and the wavelength of radiation strikes [31]. If the wavelength is smaller than the surface roughness, light is scattered diffusely. For wavelengths much larger than the roughness dimensions, the radiation is specularly reflected as from a mirror [250]. Beckmann & Spizzichino reports that reflectance is specular if the roughness ($R_a$) is smaller than one eighth ($1/8$) of the wavelength and otherwise diffuse [30]. The smaller the roughness, the higher the reflectivity: reflection from smooth and polished surfaces is mirror-like, whereas it is diffuse from rough surfaces [250]. Surfaces with roughness smaller than approximately 1.18 micrometer ($\mu$m) reflect a human's radiation (with a peak wavelength of 9.5 $\mu$m) in specular manner.

A thermal camera produces thermograms of a surface based on the incident radiation from the surface. This radiation includes the energy the surface emits (based on its temperature) as well the reflection of objects' radiation from the surrounding. If the reflectivity of all objects is diffuse the camera only views the objects in its direct field of view. However, if a surface reflects radiation in a specular manner, it acts as a mirror for the thermal camera. Thus, the camera is additionally able to view objects which are out of its direct field of view but visible through the surface's reflection. With such surfaces it is possible to extend the camera's field of view and the space of interaction, respectively.

Objects reflect thermal radiation and visual light differently. Surfaces made of different metals or with a smooth paint can act as a mirror in the thermal spectrum and can still be used for visual projection. Other materials such as transparent glass and plastic are transparent for visual light but still a mirror for thermal radiation. In the following, we show how the reflection of thermal radiation can be exploited to build interactive systems that can sense body gesture in front, besides and even behind a thermal camera. We show that a wide range of materials exist that diffuse visual light and can thus be used for projecting visual content but still reflect thermal radiation. As the human body radiation is in the F-IR range, we are interested in surfaces that have high specular reflectivity for F-IR radiation but diffuse reflectivity in the visual spectrum.

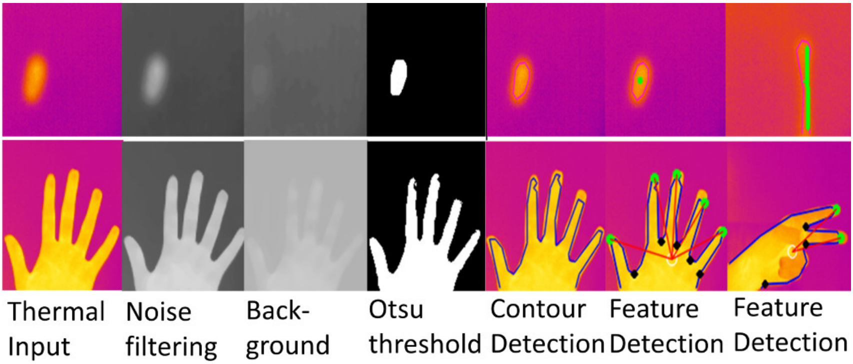



**Figure 4.8:** Feature extraction for on surface and in-air gesture.

### 4.3.3 Evaluation

We conducted a study in both indoor and outdoor settings to assess the detection accuracy of mid-air gestures performed in the reflected space using thermal reflectivity. We replicated the system reported in [210]. We recruited 30 participants (11 female, with an average age of 26 years, SD=3.8) using our university's mailing lists. All participants were students in different majors. Three participants were left handed. None of the participants had experience with thermal cameras. The participants were divided into two groups of 15; one group used the indoor setup, and the other used the outdoor setup.

#### Apparatus

The indoor and the outdoor setups were identical, including a projector and an Optris PI160 contactless thermal camera with 23° x 17° field of view(Figure 5.1). The projector was connected to a PC and displayed the tasks on a surface. The camera was mounted on a tripod and faced toward the surface from distance of 50 cm to cover the surfaces with a dimension of 30x60 cm. Since the setup was stationary no dynamic calibration was required. No special light source or illumination conditions were requisite as the thermal camera operates independent of illumination. We considered four common surfaces, also used in [210], for the experiment: glass, tile, MDF, and aluminum.

To detect mid-air gestures we implemented the algorithms described in Figure 4.8. We used OpenCV used for the hand and finger tips extraction. The frame extraction from the thermal imaging was done through the dynamic link library (DLL) the Optris camera provides. The image analysis covered the steps reported in [210] included the pre-processing, noise and background removal, and thresholding. The feature extraction, i.e. the hand and finger information, was computed from the hand contour, convex hull, and convexity defects. The mid-air gesture detection was based on a view-based approach where the fingers' positions and their relative distance were used to match the predefined gesture.

## Tasks & Procedure

To assess the detection accuracy we considered three different gestures: (1) mid-air touch interaction using one finger, (2) continuous interaction using one finger, and (3) mid-air hand postures using two fingers. We considered different tasks for each gesture. For the mid-air touch 5 points were randomly projected on the surface. Participants were asked to touch the points in-air. We considered the dragging task for the continuous interaction. A pair of points were projected and users were asked to drag an object from the start point to the end point. Users repeated this task three times. For the mid-air posture, users were asked to perform a pinch and a pan gesture. For this task, users had to use two fingers.

After welcoming participants, we described the purpose of the study. We showed them where they had to stand and where to preform the gestures. Each participant performed the tasks for four different surfaces resulting in 12 tasks (3 tasks x 4 surfaces). The order of the tasks and surfaces where randomized per participant. The light condition of the indoor setup was constant during the whole study. The outdoor setup was in a shadow for the projection to be visible and the temperature was constant (24°C). The study took approximately 30 minutes per participant. We recorded the study on video and stored the thermal video for later analysis as ground truth. The time between the camcorder and the thermal camera was synchronized. To calculate the accuracy, an experimenter watched the videos and counted the number of times a gesture was correctly recognized by the software.

| Surface | Setup | $R_a$ | Task: Touch | | Task: Continuous | | Task: Posture | | $b_{surface}$ | $T_{surface}$ | $T_c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Glass | Indoor | .004 | 97% | SD=.17 | 95% | SD=.22 | 93% | SD=.26 | 1288 | 17 | 21 |
| | Outdoor | | 96% | SD=.20 | 95% | SD=.21 | 93% | SD=.26 | | 22.7 | 24.9 |
| Tile | Indoor | .04 | 86% | SD=.35 | 88% | SD=.33 | 89% | SD=.31 | 3852 | 18 | 21 |
| | Outdoor | | 87% | SD=.34 | 88% | SD=.33 | 89% | SD=.31 | | 20.5 | 22.9 |
| MDF | Indoor | .11 | 85% | SD=.35 | 86% | SD=.33 | 88% | SD=.36 | 365 | 19 | 22 |
| | Outdoor | | 83% | SD=.37 | 86% | SD=.35 | 87% | SD=.36 | | 27.05 | 27.9 |
| Aluminum | Indoor | .33 | 56% | SD=.50 | 67% | SD=.48 | 46% | SD=.51 | 22265 | 14 | 24 |
| | Outdoor | | 58% | SD=.50 | 64% | SD=.48 | 47% | SD=.51 | | 14.07 | 24.3 |

**Table 4.3:** The recognition accuracy and the contact point temperature.

## 4.3.4 Results and Discussion

Table 4.3.3 shows the recognition accuracy for all tasks, surfaces and setups. We found very similar results for indoor and outdoor setups. Using the glass surface resulted in the highest accuracy followed by tile, MDF, and aluminum. Considering the roughness of the surfaces reported in [210] the Pearson coefficient revealed a strong inverse correlation between the accuracy and the roughness (r= -.98). Surfaces with smaller roughness result in more mirror-like reflectivity, respectively, sharper images and higher recognition accuracy. Whereas, surfaces with higher roughness have hazy reflectivity, hence, lower accuracy.

In this study, we investigated the recognition performance of the gestures performed by single and multiple fingers. However, our system could support any arbitrary gestures by feeding the tracked fingers positions to either $1 or $N gesture recognizers which recognize arbitrary gestures formed of single or multiple strokes [26, 274, 90].

## Thermal Imaging & Interactive Surfaces

The result of our study and reviewing prior work reveals that surfaces with specific properties should be used to detect interaction and create interactive surfaces using thermal imaging. We divide the interaction into two spaces: interaction on the surface through touch, and mid-air gesture interaction. In the following, we discuss which properties of surfaces should be considered to support these interactions.

## On surface interaction using heat traces

Tracking interaction on a surface using a thermal camera relies on heat traces left behind by the contact of the finger with the surface. Based on the thermodynamic laws the heat goes from the warm object to the cold object. Hence, the heat transfer between the finger and surface occurs as far as their temperature differs. The amount and direction of heat transferred principally relies on the surface's material property, known as the thermal contact conductance [49]. The thermal contact conductance refers to the conductivity of heat between two objects in contact. The amount of heat transferred (conducted) between the hand and the surface in contact could be either reflected or absorbed by the surface.

To determine on-surface interaction, we are interested in the heat trace and the temperature change at the point of contact. Ray suggests a simple model that calculates the temperature at the contact point [207]. In our case, the contact point is between the user's skin and the surface. Hence the temperature at the contact point ($T_c$) in $^\circ C$ is as follows:

$$T_c = \frac{b_{skin} T_{skin} + b_{surface} T_{surface}}{b_{skin} + b_{surface}} \tag{4.1}$$

$$b = \sqrt{K.P.C} \tag{4.2}$$

The $T_c$ depends on the temperature of the two contact points ($T_{skin}$ and $T_{surface}$) as well as their thermal penetration coefficient ($b$). The $b$ depicts the amount of heat penetrated and absorbed by a surface. It is expressed in terms of thermal conductivity (K), thermal density (P), and specific heat capacity (C) [190] (Equation 4.2). The $b$ of human skin for short contact is 1000 $JS^{-1/2}m^{-2}K^{-1}$ [190]. On the other hand, the detection of temperature changes at the contact point depends on the camera's sensitivity. The changes must be higher than the camera's temperature sensitivity to be visible by the camera.

To detect heat traces, it is necessary to consider its decay time. Based on the Netwon's law of cooling, the rate of heat loss of a body ($Rate_{cooling}$) is proportional to the temperature difference between the body and its surrounding. The higher the difference is, the lower is the cooling rate, thus, the trace lasts longer. If the cooling rate ($Rate_{cooling}$) is smaller than the time one frame lasts ($1/FrameRate_{camera}$), the camera can not sample the trace before it decays. It should be mentioned that the cooling rate depends on other additional factors such as the surface area of the heat being transferred and the heat transfer coefficient between surfaces.

We calculated $T_c$ for the surfaces used in the study by measuring the $T_{skin}$, the $T_{surface}$, and obtaining the $b_{surface}$ from [190] (Table 4.3.3). The sensitivity of the camera used is $0.08°C$ and it's frame rate is 120Hz. Based on our camera property, the change in temperature at the contact point should be bigger than $0.08°C$ and should last at least 8.3 milliseconds. The result unveiled that the difference between $T_c$ and $T_{surface}$ is more than threshold for all surfaces except for the aluminum. Thus, the temperature changes on an aluminum surface is invisible to our the camera. In [210], it is empirically tested and reported that no trace can be detected on the aluminum using this thermal camera. A camera with higher sensitivity and/or a higher frame rate may reveal other results.

## Mid-air gestures through thermal reflectivity

Previous work discussed how the camera's FOV can be extend through thermal reflection [210]. This allows one to observe in-air interaction in the direct FOV as well as behind the camera. It has been discussed that the thermal reflectivity should be specular and depends on the roughness of a surface. Based on the thermal radiation of human body, it is reported that the roughness ($R_a$) of the surface should not exceed 1/8 of the human body radiation, i.e. 1.18 micrometer, to recognize interaction behind the camera's direct FOV. The results of our study reveal that the lower the roughness the sharper the rendered image from the reflection, respectively, higher is the recognition performance. Otherwise, the interaction in extended space is too blurry and cannot be used.

## Material Space for Interactive Surfaces

Knowing certain information about a surface enables us to determine if on-surface interaction and interaction through thermal reflection can be used. To support touch interaction on a surface, we need the thermal penetration coefficient ($b$) of the surface to calculate the temperature at the contact point ($T_c$). The difference between $T_c$ and the environments plays an important role to determine how fast a trace decays. The higher the difference the longer the trace lasts. When $T_c$ is higher than the camera's sensitivity and the trace lasts at least for one frame, it is possible to detect interaction on the surface. Based on our setup and our thermal camera, surfaces such as glass, MDF, tile, etc. can be used for touch interaction on the surface. Aluminum cannot be used as the trace decays faster than one frame lasts (8 ms). Figure 4.9 shows the material space for an Optris PI160.

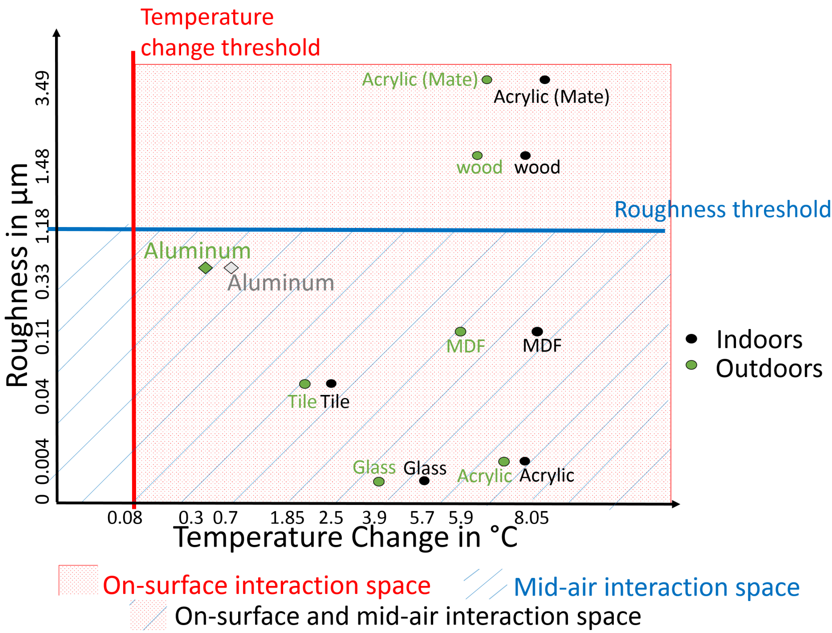



**Figure 4.9:** Material space for Optris PI160-Based Interaction

A surface with roughness below 1.18 micrometer is required for using thermal reflectivity to extend the mid-air interaction space (Figure 4.9). For our setup, surfaces such as glass, aluminum and MDF have this property, but wood or matte acrylic cannot be used to extend the mid-air interaction space. Furthermore, the higher the roughness is the lower is the recognition accuracy of interaction in the extended space. For example, MDF results in a lower accuracy than tile.

### 4.3.5 Summary

Thermal imaging has shown promising potential for interactive systems. However, there are still limitations and constraints concerning deploying thermal cameras as sensing technologies. Sensitivity of the camera and its frame rate are the main constraints. These affect the sharpness of reflected images rendered. The resolution of the camera should be also considered. We report the material space (Figure 4.9) for a specific thermal camera (Optris PI160). Using the approach described above, it can be easily derived for cameras with improved properties.

In this work we assessed the recognition accuracy of mid-air interactions sensed through thermal reflectivity using surfaces with different reflection characteristics in indoor and outdoor setups. We further derived the material space and the constraints of interaction enabled through thermal imaging. The guideline allows identifying whether a surface supports on-surface interactions and/or mid-air gestures using a thermal camera as the sensing technology. Knowing the thermal penetration coefficient of a surface, it is possible to determine if heat traces last long enough at the contact point on the surface to be detected by the thermal camera. Further, it is possible to find out if the mid-air interaction space can be extended beyond the camera's direct FOV through thermal reflectivity by knowing the roughness of the surface.

Our work as well as previous work that uses thermal imaging for interactive systems uses standard computer vision techniques originally developed for the visual spectrum. This approach already provides reasonable performance. However, significant improvements can be expected when using algorithms specifically designed to exploit the characteristics of thermal imaging. To sense users through thermal reflection techniques that were initially designed to remove reflections from recorded thermal images [250] could be applied. By separating the scene sensed through reflection and the directly observed, both scenes could be analyzed independently. A further improvement could be achieved by using more advanced thermal cameras. While such cameras were once only available for the military, they are now also becoming available for normal use.

## 4.4 ThermalAttacks: Detecting PINs and Patterns

*This section is based on the following publication:*

- Y. Abdelrahman, M. Khamis, S. Schneegass, and F. Alt. Stay cool! understanding thermal attacks on mobile-based user authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3751–3763. ACM, 2017

Threats and privacy concerns emerged as an important topic, brought up from our investigation of the users' perception of thermal imaging (cf. Section 3.2). Chapter 4 and 5, showed the opportunities of having amplified perception. As imaging technology, we can imagine a scenario where technology misuse is exercised.

While in previous sections we utilized thermal conductance to recognize touch points for amplified interaction space, in this work we investigate its reliability to infer passwords from heat traces left on touch screens after authentication.

The degree to which the user's privacy is threatened will determine how accepted and acceptable thermal imaging will be in the future, as privacy concerns are known to influence users' technology use decisions [198]. Thus, users need to be aware of the potential threats and means to overcome them. In this section, we investigate thermal attacks against PINs and patterns on mobile devices. We studied *Thermal Attacks*, their feasibility, and potential means and guidelines to overcome them.

Thermal cameras allow thermal attacks, where heat traces resulting from authentication can be used to reconstruct passwords, presenting a new threat to user privacy on mobile devices. In this work we investigate the viability of exploiting thermal imaging to infer PINs and patterns on mobile devices and evaluate how the properties of PINs and patterns influence their thermal attack resistance. We conclude by recommendations for users and designers of authentication schemes on how to resist thermal attacks.

### 4.4.1 Related Work

Our work builds on two strands of prior work: (1) thermal imaging and (2) the different types of threats to user authentication on mobile devices.

There are multiple differences between properties of thermal imaging and those of visible light. The first thermal property is heat radiation. Compared to visible light, heat radiation has different reflection properties that depend on the surface [8]. Thermal reflections were exploited in previous work to enable body-worn and hand-held devices to detect mid-air gestures [210].

The second unique property is that thermal imaging is independent of light and coloring conditions, which allows thermal cameras to be used for face and expression recognition [134, 138]. Thermal cameras can provide information about the sensed body's temperature, which can be used to infer the physiological and cognitive state of users in a contact free manner [208] by, for example, evaluating their stress levels [131].

A third unique property is that thermal imaging is capable of detecting input that has been performed in the past. When a user touches a point on a surface, heat is transferred from the user to the surface, generating heat traces that slowly fade away. These traces can be detected using thermal imaging. Heat traces have been utilized for input [86, 150, 210] and to authenticate users based on their thermal hand print [47].

In this work, we investigate the use of thermal imaging to infer passwords entered on mobile devices, exploiting the fact that heat traces take time to fade away. We investigate the thermal properties of state-of-the-art touch screens and study the impact of password properties on the heat trace and, thus, the successful retrieval of passwords via thermal imaging.

#### Threats to Authentication on Mobile Devices

Mobile devices, such as tablets and smart phones, allow access to a plethora of private content. Prior work investigated a number of threat models that put the user's private data at risk.

One of the most discussed threats is shoulder surfing attacks, in which an observer attempts to eavesdrop a user to uncover private information, among which are login credentials [66]. Different approaches have been introduced to mitigate shoulder surfing attacks, ranging from adding random cues [33, 34, 35, 251], splitting the attackers' attention by requiring them to observe multiple cues [57, 133], and disguising the user input [58, 96]. In addition to focusing on login credentials, research has also investigated methods to protect users from shoulder surfing text messages [67] and pictures [255]. Most of the schemes that counter shoulder surfing address a threat model where the attacker can clearly observe the password entry once. Other threat models cover multiple observation attacks [103, 133, 184, 265] or video attacks [57, 251].

Another type of attack that has been addressed by previous work is smudge attack, in which an attacker exploits the oily residues left on the touch screen after interaction to uncover the password [28]. Smudge attacks perform particularly well against patterns, as smudges give hints on where the pattern started. However they can hardly provide any useful information about the order of PIN entries. Approaches to mitigate smudge attacks include graphically transforming the visual cue on which the password is entered [220, 256], introducing a random element that leads to different smudges at every authentication attempt [256], or using multiple fingers to increase the complexity of the pattern [183]. Threat models that consider smudge attacks assume that the attacker has access to the mobile device, in addition to clearly visible smudge traces and optimal lighting conditions to see the smudges clearly.

### 4.4.2 Thermal Attacks

Thermal image attacks exploit properties of thermal imaging. Namely, heat traces are transferred from the user's hands to the touch screen during authentication. These traces fade away slowly [150], allowing thermal cameras to perceive which parts of the display have been touched even after the user had already entered the password. Similar to shoulder surfing, thermal attacks leak information about the order of entered PINs and patterns [28]. In contrast to shoulder surfing, however, thermal attacks can be performed after the user has left the device. This gives attackers an advantage as they no longer need to observe the user while authenticating, which makes the attack more subtle and eliminates hand occlusions. Although thermal images can be distorted by interaction, a user who performs limited interactions or leaves the device after authentication is still vulnerable to thermal attacks.

Mowery et al. investigated the effectiveness of thermal attacks on ATMs with plastic keypads [174]. They found that thermal attacks are feasible even after the user authenticated. While Mowery et al. investigated thermal attacks on plastic keypads of ATMs, little work was done regarding thermal attacks on mobile devices and other touch screens devices. In a preliminary study, Andriotis et al. [24] were able to observe heat traces resulting from entering a pattern for 3 seconds after authentication. This allowed them to retrieve parts of the pattern.

We perform an in-depth analysis of how well thermal attacks perform on PINs and patterns on mobile device touch screens with respect to different password properties. We also consider duplicate digits in PINs, and overlaps in patterns. To do this, we implemented ThermalAnalyzer, which automatically retrieves passwords from heat traces. ThermalAnalyzershows that thermal attacks can be successful even if they take place 30 seconds after authentication (i.e. 10 times longer compared to previous work [24]).

### 4.4.3 Understanding thermal attacks

Our work relies on the phenomenon of heat transfer from one object to another. Heat transfers from users' hands to surfaces they interact with, leaving traces behind that can be analyzed. This relies on the surface's material property know as thermal contact conductance [49], which refers to the conductivity of heat between two objects (surfaces) that are in contact.

According to the black body model [113], any object above absolute zero (e.g. surrounding objects in our environment) emits thermal radiation. This radiation is absorbed, reflected, and transmitted. However, for fully opaque surfaces the transmitted portion is discarded [81]. This limits the portions of interest to the reflected and absorbed radiation. Hence, thermal radiation could be presented as in $Thermal\ reflectivity + Thermal\ absorptivity = 1$.

As soon as an object contacts a surface, thermal radiation is transmitted and absorbed by the surface, causing a temperature change. This leads to heat traces accumulating on the surface. To compute the transferred heat and identify whether or not it is detectable by commercial thermal cameras, we measured the temperature at the contact point ($T_{contact}$). We used a well-established model by Ray [207] to compute the temperature at the contact point of the two bodies. In our scenario, the two bodies are the human skin (i.e. the user's finger), and the mobile device's touchscreen (i.e. a plate of Gorilla glass).

$$T_{contact} = \frac{b_{skin}T_{skin} + b_{gorilla\ glass}T_{gorilla\ glass}}{b_{skin} + b_{gorilla\ glass}} \tag{4.3}$$

$$b = \sqrt{K.P.C} \tag{4.4}$$

$T_{contact}$ depends on the temperature of the contact points ($T_{skin}$ and $T_{gorillaglass}$) as well as their *thermal penetration coefficient (b)*. It is the amount of thermal energy penetrated and absorbed by the surface. The $b$ is defined in Equation 4.4. It is composed of the product of thermal conductivity (K), thermal density (P), and specific heat capacity (C) [190]. The $b$ of human skin and the gorilla glass for short contact are 1000 $JS^{-1/2}m^{-2}K^{-1}$ [190] and 1385 $JS^{-1/2}m^{-2}K^{-1}$ [241] [18] respectively.

Additionally, the detection of temperature changes at the contact point depends on the camera's sensitivity. The change in temperature must be higher than the camera's temperature sensitivity to be distinguishable by the camera. For example, if the touch screen's glass has a temperature $T_{gorillaglass}$ of 23°C and the user's hand temperature $T_{user}$ is 30°C, then $T_{contact}$ would be 25.9°C according to Equation 4.3. This results in a temperature difference of 2.9°C ($T_{contact} - T_{gorillaglass}$). Hence, a thermal camera with thermal sensitivity $\leq$ 2.9°C would be able to recover the order in which a PIN/pattern entry was performed by utilizing the heat trace decays. In our work, the thermal camera has a thermal sensitivity of 0.04°C, allowing different hand temperatures to be sensed.

## Threat Model

In our threat model, the attacker (i.e. a person who wants to access a device without permission) waits for the victim to complete the authentication process and to leave the mobile device. This could be the case when the user quickly checks his latest messages before getting something to drink from the coffee machine, while leaving the device on his/her desk. To ensure optimal conditions for the attacker in our threat model, the user does not interact with the device but merely authenticates (e.g. to check an update from a notification or a widget) then leaves the device idle. The attacker then uses a thermal camera (e.g., integrated into a smart phone) to take a thermal image of the device's touchscreen. The attacker then analyzes the thermal image in a manner similar to our analysis presented in the following section to identify the PIN/pattern. Similar to previously discussed threat models [103, 184, 133, 265], the attacker exploits a chance where the device is unattended to login and access the user's private information.

[18] This value was confirmed by lab measurements by the Institute of Applied Optics in our university

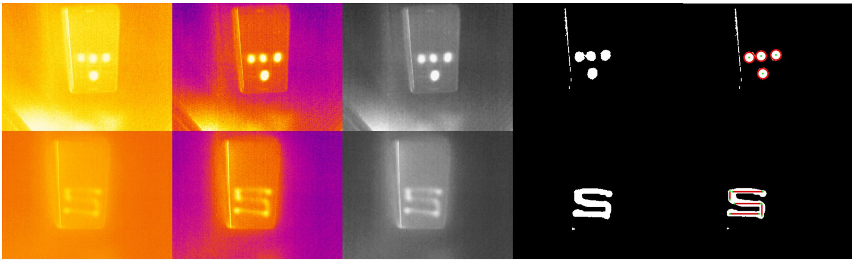



**Figure 4.10:** Recognition pipeline of PINs (top) and patterns (bottom).

### 4.4.4 ThermalAnalyzer

In the following, we describe the design and implementation of the ThermalAnalyzer. The ThermalAnalyzerconsists of a thermal camera capturing an image and a recognition pipeline used to extract the PINs and patterns.

## Recognition Pipeline

The recognition pipeline consists of six steps, performed to extract a PIN or patternsfrom an image (Figure 5.2). The steps are performed using OpenCV[19] and include frame extraction, pre-processing, noise and background removal, and thresholding. The final step is feature extraction to deduce the touch points' location and temperature information.

**Frame Extraction and Camera Configuration** We captured the thermal image through the Optris thermal camera API[20]. Using the interprocessor communication, we capture the frames in a 16-bit color format along with the encoding of the temperature information. We configured the camera using its API to capture temperature values between 19°C to 32°C. This was done to achieve higher contrast of colors that represent different temperature values as depicted in Figure 5.2. For each captured frame a pre-processing procedure is performed. This included noise filtering, background subtraction, and thresholding.

**Noise Filtering** We adapted the noise filtering process used by [8, 150, 210] by applying a 5×5 px median filter, converting the image to grey scale and reapplying the filter for enhanced noise reduction.

[19] `http://opencv.org/`

[20] `http://www.optris.com/software`

**Background Subtraction** We built a semi-static background model for background subtraction. A static model is preferred in our case as we want the heat trace detected to last over the frames and not to be adsorbed by a dynamic background model. Yet, on the other hand a dynamic model is required to tolerate slight temperature difference of the device along the operation. Hence, we built a semi-static background model, where the update is controlled by the learning rate ($\alpha$) parameter, which is a value that controls the rate of background model updates. An $\alpha$ value of 0.001 showed the best result in preliminary tests. As a result, the latest heat trace stemming from password entry stays in the foreground, whereas heat traces from slight changes from the environment temperature are merged with the background.

**Thresholding** To segment the regions that are relevant to identifying heat traces (Figure 5.2), we used Otsu's thresholding method [187]. The frame is classified into two sets of pixels with minimum overlap between them based on a dynamically computed threshold by Otsu's algorithm. Then, we applied an additional morphological closing operation to highlight the boundaries of the thresholded foreground and reduce the background.

**Feature Extraction** Our features are classified into (1) circular fitted traces for PIN detection and (2) line fitted traces for patternsdetection.

The heat trace is detectable via extracting the contours from the binary images, where the image is scanned to detect arrays of contours. Similar to the work of Sahami et al. [210], we used a *circular fitted* contour detection to identify the PIN entries. The contour center is computed as the spatial moment of the extracted contour. Using the same approach for detecting the circular fitted heat traces, we used the Hough Transform [65] for extracting *line fitted* contour detection to identify the entered patterns, as depicted in Figure 5.2.

**PIN and Pattern Sequence Detection** At this point in the processing pipeline, the PIN or patternentry has been extracted from the captured frames but with no information about the sequence of entry. To infer the sequence for the PINs, we utilized a pre-set frame with the keypad to identify the PIN location using squares. The squares represent the entire set of regions of interest (ROI). Mowery et al. [174] reported that representing the ROI with the mean temperature yields best performance for recovering the order of the entry sequence. Hence, we compute the mean temperature for each ROI, and sort them based on their weights.

To identify duplicate entries, we compute the overall average temperature of each digit. Thereby, the background temperature is subtracted. Hence, a digit that was never pressed would have a temperature value of almost zero. Consequently, duplicate entries (i.e. the digit that was touched multiple times) have a value that exceeds the overall average. The number of duplicates can be inferred from the relative temperature values of the overall number of detected presses. In summary, given a four-digit PIN, there would be four detection scenarios:

- Four different heat traces: This means there are no duplicates. Hence, ordering the traces based on their temperature in a descending manner would infer the sequence.
- Three different heat traces: The heat trace that has a temperature of $T_{contact}$ is the last entry in the PIN, as it will maintain the $T_{contact}$ value. This leaves only 3 possibilities for the remaining sequence, which are sufficient for the attacker to try without being locked out. This approach, however, will work with recently captured frames as the heat trace, i.e. $T_{contact}$, decays over time.
- Two different heat traces: According to the relative ratios of the weights, the number of repetitions of each digit is identified. Normalizing the weights would then show the last touched digit. Once the last digit is identified, the attacker can tell whether it is the duplicated one (i.e. the other duplicate is either in position 1, 2 or 3, while the remaining digits are ordered according to their heat traces), or the last digit is a non-duplicated digit, hence the attacker has only 3 possibilities to try without being locked out.
- One heat trace: This means that the PIN consists of the same number repeated 4 times.

One of the former three conditions could be experienced due to heat trace decay. In that case, we identified the missing digit to be unidentified and set it to be the beginning of the PIN (e.g., if 3 traces where detected with no evidence of duplicates, the first digit is labeled unknown and the remaining three are sorted by their temperature weights).

The same approach was followed for the patterns, where the extracted lines are analyzed and ordered by their mean temperature. Additionally, the temperatures of the tips of the extracted lines were compared to identify the direction. Our algorithm does not account for a specific patternlength, hence we present the available heat trace to be the regenerated pattern.

For a more conservative analysis, ThermalAnalyzeris not optimized for detecting patternsof specific length (max of 9). This is because in our threat model, and most likely in a real scenario, the attacker does not know the pattern length. This means that in cases where the ThermalAnalyzergenerated a guessthat is of length *n* instead of 9, the heat traces of the remaining 9–n had already decayed by the time of the attack.

### 4.4.5 Evaluation

Despite the variety of authentication schemes that were introduced in past years, personal identification numbers (PINs) are one of the most commonly used schemes [254]. Moreover, as Android devices dominate the market, there is an increasing adoption of patterns, which is an Android graphical password scheme where users draw a line pattern that connects dots displayed in a $3 \times 3$ grid [242].

The aim of this study is to analyze thermal images of a smart phone screen after a user has entered a password. These images are evaluated using ThermalAnalyzer. We particularly focus on understanding how (1) different authentication schemes (2) the time between password entry and attack, and (3) different password properties influence the feasibility of thermal attacks.

#### Design

The study uses a repeated measures design, where all participants were exposed to all conditions. We studied the effect of three independent variables on the success of thermal attacks: (1) the password type: whether the used scheme is PIN or pattern, (2) the age of the heat trace: we recorded the heat traces continuously for 60 seconds to investigate for how long they remained exploitable by an attacker, and (3) the properties of the PINs and patterns.

In case of PIN, the property we studied was the number of duplicates in the PIN. On one hand duplicates distort the heat traces, making the differences between the first entry and the last entry less distinguishable. On the other hand the presence of duplicates reduces the password space, which means that less information from the thermal attack would be sufficient to uncover the password. We studied the influence of having No-duplicates, 1-duplicate, and 2-duplicates(e.g., 1234, 1233, and 1222 respectively).

In case of patterns, we investigated the effect of the number of overlaps [253] in the pattern. An overlap occurs when the user's finger passes through a node that is already selected. We expect that overlaps can distort the heat traces enough to

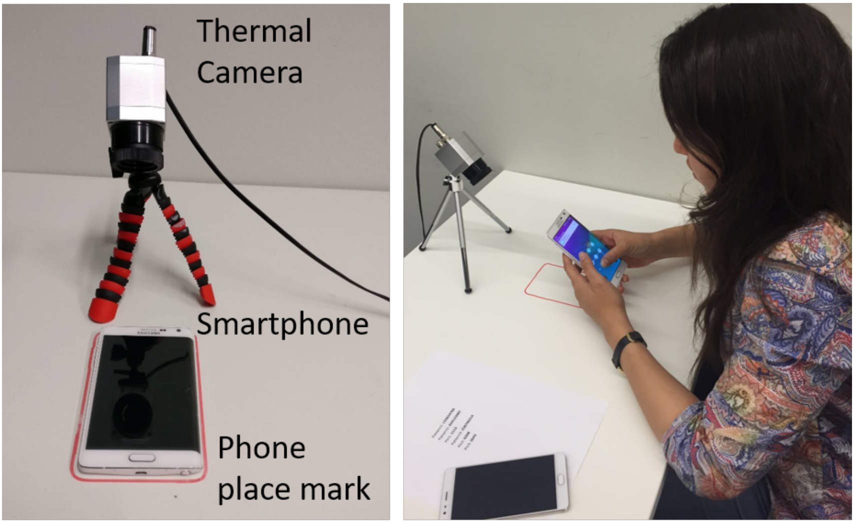



**Figure 4.11:** Setup with the thermal camera capturing the phone's screen.

make it infeasible to reconstruct the entered pattern. We studied the influence of having one, two, or no overlapping lines in the patterns.

## Apparatus

Our setup included two Samsung Galaxy Note Edge smartphones, a thermal camera (Optris PI450 [21]), and an RGB camera GoPro Hero3 each mounted on a tripod. One smartphone was used for practicing the passwords and the other one for the actual input. The thermal camera has an optical resolution of 382 x 288 pixels and a frame rate of 80Hz. It is able to measure temperatures between -20°C and 900°C, and operates with a thermal sensitivity of 0.04°C represented by the noise equivalent temperature difference (NETD)[22]. The wavelengths captured by the camera are in the spectral range between 7.5$\mu$m and 13$\mu$m. The lens we used provides a 80°$\times$ 58°field of view. The thermal camera uses USB as power source as well as to transfer data. It provides temperature information in the form

[21] http://www.optris.com/thermal-imager-pi400

[22] NETD refers to the electronic noise that is interpreted as a temperature difference of an object

of 16-bit color values encoding the temperature information. The marker in front of the camera (cf., Figure 5.1) indicates the optimal position of the smart phone to video-tape the heat trails with the thermal camera while minimizing thermal reflection. Additionally we recorded the whole study using an RGB video camera. The RGB video feed was later used to determine the time at which the users fingers were no longer in contact with the screen.

## Participant and Procedure

We recruited 18 participants (10 female, 8 male) with an average age of 28.3 years ($SD = 4.7$) using our University's mailing lists. All participants were students in different majors. Two participants were left handed. None of the participants had any previous experience with thermal cameras. After the participants arrived in the lab, we first asked them to sign a consent form and explained the purpose of the study. Next, we handed out a set of PINs and patternsprinted on cards as well as the two smartphones to the participants. To avoid errors and pauses during entry, we asked the participants to familiarize themselves beforehand with the passwords by entering them multiple times on the practice smartphone first. We instructed the participants to enter the password, then immediately place the study smartphone on a place mark on the table in front of them (cf., Figure 5.1). We waited for three minutes between each entry, to ensure full heat trace decay of the previous entry. Each participant entered three passwords of each type (i.e., 18 passwords). The order was counterbalanced using a Latin-square.

The study took approximately 40 minutes per participant We recorded the whole study on video for post-hoc analysis of the password input times. Throughout the experiment we recorded the temperature of the participant's dominant hand (i.e. the hand used to enter the password), in addition to the phone's temperature. The experiment was conducted in a maintained room temperature of 24°C.

To analyze the thermal attacks, we considered two approaches: (1) visually inspecting heat traces and (2) using our computer vision approach ThermalAnalyzer. The visual analysis was done by one of the authors who was not aware and had not been exposed to the list of entered passwords. Additionally, the feed was analyzed from the thermal cameras. Using the approach explained earlier, the author reported the regenerated PINs and patternsin csv file defining all possible PIN combinations in cases where there were duplicates.

## 4.4.6 Results

To evaluate the success of thermal attacks against PINs and patterns, we measured

1. The success rate: the percentage of cases in which the thermal attack successfully revealed the entire password correctly.
2. The Levenshtein distance: the distance between the generated guesses and the correct password.

The success rate and Levenshtein distance were used in previous work to reflect how successful attacks are (success rate) and how close the guess is to the original password (Levenshtein distance) [59, 133, 252]. We only considered the PINs and patternsfrom the ThermalAnalyzer. We investigated the effect of three independent variables: (1) authentication scheme, (2) age of the heat trace and (3) password properties. The tasks performed during the study typically require between 26% to 44% CPU usage.

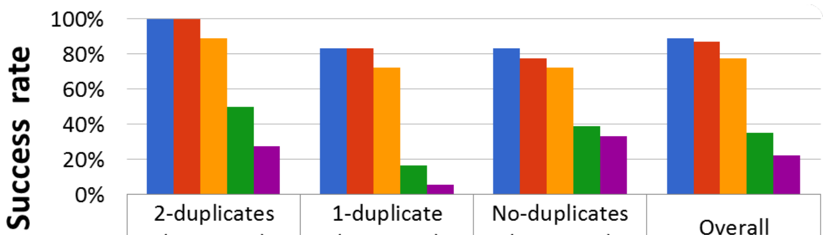
## Statistical Analysis

As we have three independent variables, we analyzed the data using a three-factor repeated measures ANOVA (with Greenhouse-Geisser correction if sphericity was violated). This was followed by post-hoc pairwise comparisons using Bonferroni-corrected t-tests.

Figures 4.12 and 4.14 show the success rate per age of the heat traces and password property. Additionally, Figures 4.13 and 4.15 show the Levenshtein distances per age of the heat traces and password property. The results show that thermal attacks are more successful against PINs than against patterns.

| | **PIN** | | **Pattern** | |
|---|---|---|---|---|
| **Age** | **Levenshtein** | **Success Rate** | **Levenshtein** | **Success Rate** |
| immediate | M=0.222, SD=0.76 | M=0.89, SD=0.08 | M=0.222, SD=0.76 | M=0.46, SD=0.40 |
| 15 seconds trace | M=0.222, SD=0.76 | M=0.87, SD=0.09 | M=0.315, SD=0.139 | M=0.44, SD=0.40 |
| 30 seconds trace | M=0.407, SD=0.134 | M=0.78, SD=0.08 | M=0.407, SD=0.134 | (M = 1, SD = 0.39)0.44 |
| 45 seconds trace | M=1.39, SD=0.2 | M=0.35, SD=0.14 | M=1.39, SD=0.2 | M=0.20, SD=0.14 |
| 60 seconds trace | M=1.94, SD=0.23 | M=0.22, SD=0.12 | M=3.8, SD=0.32 | M=0.11, SD=0.09 |

**Table 4.4:** Success rate and Levenshtein distances for different ages.




| | 2-duplicates (e.g. 1222) | 1-duplicate (e.g. 1233) | No-duplicates (e.g. 1234) | Overall |
|---|---|---|---|---|
| Immediate | 100% | 83% | 83% | 89% |
| 15 seconds | 100% | 83% | 78% | 87% |
| 30 seconds | 89% | 72% | 72% | 78% |
| 45 seconds | 50% | 17% | 39% | 35% |
| 60 seconds | 28% | 6% | 33% | 22% |

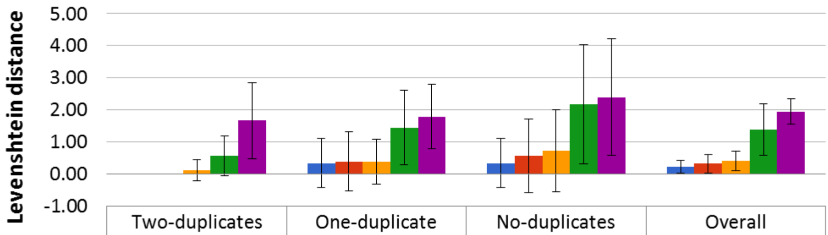
**Figure 4.12:** The success rate of thermal attack against PINs.

## Authentication Scheme: PINs vs Patterns

Overall, thermal attacks were more successful for PINs (M = 0.62, SD = 0.31) than for patterns (M = 0.32, SD = 0.16). Similarly, the Levenshtein is shorter for PINs (M = 0.856, SD = 0.127) than for patterns (M = 3.14, SD = 0.28). We found a significant main effect of password type on the Levenshtein distance between the guess and the entered password $F_{1,17} = 91.923$, $p < 0.001$. Post-hoc analysis showed significant differences $(p < 0.001)$ between passwords of type PIN (M = 0.856, SD = 0.127) compared to those of type pattern (M = 3.14, SD = 0.28). This means guesses against PINs are generally closer to the original password compared to those against patterns.

## Age of Heat Traces

**PINs** Looking at the age of the heat trace, the results show that the earlier the heat attack is performed, the higher the success rate and the lower the Levenshtein distances are (cf., Table 4.4.6). The results of the ANOVA revealed a significant main effect of the heat trace's age on the Levenshtein distance between the correct password and the guess $F_{1.79,30.45} = 41.7$, $p < 0.001$. Post-hoc analysis using Bonferroni corrected t-tests showed statistically significant differences between 60 seconds and all other durations $(p < 0.001)$ as well as between 45 seconds



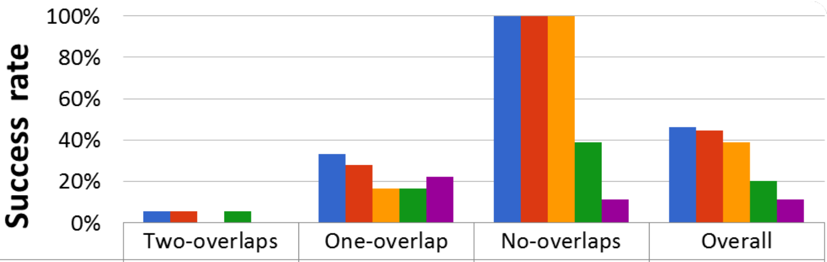

| | Two-duplicates | One-duplicate | No-duplicates | Overall |
|---|---|---|---|---|
| immediate | 0.00 | 0.33 | 0.33 | 0.22 |
| 15 sec | 0.00 | 0.39 | 0.56 | 0.31 |
| 30 sec | 0.11 | 0.39 | 0.72 | 0.41 |
| 45 sec | 0.56 | 1.44 | 2.17 | 1.39 |
| 60 sec | 1.67 | 1.78 | 2.39 | 1.94 |

**Figure 4.13:** The mean and standard deviation Levenshtein distances.

and all other durations ($p < 0.001$). This shows that thermal attacks against PINs that take place within the first 30 seconds after authentication result in guessesthat are significantly closer to the correct password compared to those done after 30 seconds. This is also reflected in the success rate as shown in Figure 4.12. Overall, this suggests that thermal attacks are very effective against PINs when performed within 30 seconds after authentication.

**Patterns** Similar to the results of the PINs, the results for the patternsshow that the older the traces are, the less likely a thermal attack is successful and the higher the Levenshtein distances are (cf. Table 4.4.6). We found a significant main effect of the heat trace's age on the Levenshtein distance between the correct patternand the guess $F_{2.228,38.876} = 13.295$, $p < 0.001$. Post-hoc analysis using Bonferroni corrected t-tests showed significant differences ($p < 0.05$) between 60 seconds and all other durations.

This shows that thermal attacks against patternsthat take place 60 seconds after authentication result in guesses that are significantly farther away from the correct password, compared to those done within the first 45 seconds. This is also reflected in the success rate shown in Figure 4.14. Overall, this suggests that thermal attacks are very effective against patternswhen performed within 45 seconds after authentication.




| | Two-overlaps | One-overlap | No-overlaps | Overall |
|---|---|---|---|---|
| immediate | 5.56% | 33.33% | 100.00% | 46.30% |
| 15 sec | 5.56% | 27.78% | 100.00% | 44.44% |
| 30 sec | 0.00% | 16.67% | 100.00% | 38.89% |
| 45 sec | 5.56% | 16.67% | 38.89% | 20.37% |
| 60 sec | 0.00% | 22.22% | 11.11% | 11.11% |

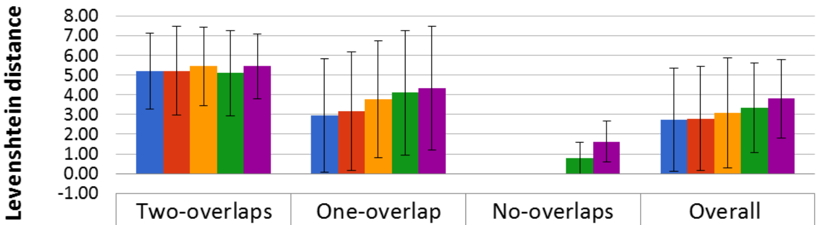
**Figure 4.14:** The success rate of thermal attack against patterns.

## Hand and Screen Temperature

We found that the difference in temperature ($D_t$) between the hand and screen influences the success of a thermal attack. The higher $D_t$, the more successful is a thermal attack, since more thermal energy is transferred to the screen (cf. Equation 4.3). Using Pearson's product-moment correlation, we found that the correlation between $D_t$ and the successful thermal attack rates increases from 0.55 (at 0 seconds) to 0.85 (at 60 seconds). This means that there is a strong correlation between $D_t$ and the success of an attack and that $D_t$ is particularly important for attacks happening some time after authentication.

## Password Properties

**PINs Duplicates** We found a significant main effect of number of duplicate digits on resistance to thermal attacks $F_{2,34} = 13.23$, $p < 0.01$. Post-hoc analysis revealed statistically significant differences ($p < 0.05$) between No-duplicates (M = 1.23, SD = 0.25) and 2-duplicates (M = 0.47, SD = 0.08) and between 1-duplicate (M = 0.87, SD = 0.15) and 2-duplicates (M = 0.47, SD = 0.08). This means that the more duplicates a PIN has, the closer the guessis to the correct PIN.




| | Two-overlaps | One-overlap | No-overlaps | Overall |
|---|---|---|---|---|
| immediate | 5.22 | 2.94 | 0.00 | 2.72 |
| 15 sec | 5.22 | 3.17 | 0.00 | 2.80 |
| 30 sec | 5.44 | 3.78 | 0.00 | 3.07 |
| 45 sec | 5.11 | 4.11 | 0.78 | 3.33 |
| 60 sec | 5.44 | 4.33 | 1.61 | 3.80 |

**Figure 4.15:** The mean and standard deviation Levenshtein.

This shows that although the presence of duplicate digits makes it harder to determine the order of the detected touches, the approach is able to determine if a digit is repeated two or three times. As a result, the security added by overwritten heat traces in case of duplicate PINs is outweighed by the significantly reduced password space.

**Patterns Overlaps** We found a significant main effect of the number of overlaps on the distance between the correct patternand the guess $F_{1.441,24.503} = 28.563$, $p < 0.001$. Post-hoc analysis showed significant differences between two pairs ($p < 0.001$): patterns with no overlaps (M = 0.48, SD = 0.08) compared to those with one overlap (M = 3.67, SD = 0.68), and between patterns with no overlaps (M = 0.478, SD = 0.08) compared to those with two-overlaps (M = 5.29, SD = 0.43). No significant differences were found between the third pair ($p > 0.05$).

This shows that although patternscan be successfully uncovered using thermal attacks up to 30 seconds after authentication (100% success rate), the presence of overlaps significantly increases its resistance against thermal attacks.

## Recommendations to Resist Thermal Attacks

There are ways to resist thermal attacks. We present three categories: (1) based on the results of our study, we are able to guide users in selecting PINs/patterns that are resistant to thermal attacks, (2) based on a literature review, we recommend schemes that are theoretically unaffected by thermal attacks, and (3) we present novel approaches that distort the heat traces, reducing the chances for successful thermal attacks.

**Selection of PINs and Patterns** Our results indicate that adding a single overlap in an authentication pattern significantly increases the resistance to thermal attacks. When it comes to PINs, although duplicates distort the heat traces thermal attacks rely on, other factors also contribute to the ease/difficulty of uncovering duplicate PINs.

We recommend to increase the resistance of PINs against thermal attacks by increasing the number of digits in the PIN. The longer the PIN the longer it takes the user to enter it, which would in turn decrease the intensity of heat traces of the first digits by the time the user authenticates.

**Thermal Attack Resilient Schemes** Many authentication schemes have been proposed to resist different types of attacks. We are not aware of systems built with the main aim of resisting thermal attacks on touch screens. However, some existing knowledge-based schemes do resist them by design.

One group of authentication schemes resilient against thermal attacks relies on one or more modalities other than touch input. For example, biometrics schemes (for example, [42, 56, 109, 111]) rely on data collected by sensors, such as accelerometers, to identify the user. Since they do not use the touch screen for dedicated input, they are not vulnerable to thermal attacks.

Similarly, authentication schemes that combine touch input with another modalites increase the resilience towards thermal attacks. PhoneLock [33], SpinLock [34], TimeLock [35], and ColorLock [35] augment PIN entry by using auditory and haptic cues the user needs to respond to when authenticating. These cues are randomized to counter shoulder-surfing attacks. Other examples utilize eye movements. For example Liu et al. [157] and Bulling et al. [41] used gaze input to authenticate. Similarly, Khamis et al. [133] introduced GazeTouchPass which combines gaze gestures and touch-input. Depending on the authentication scheme, the use of thermal cameras can still help the attacker to reveal the part of the input made on the touch screen. Being untied to the touchscreen, thermal attacks against these schemes would fail to uncover the PIN.

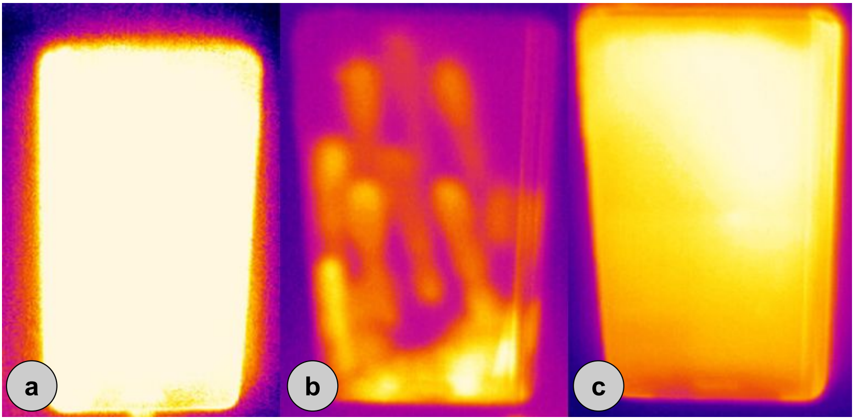



**Figure 4.16:** Approaches for resisting thermal attacks.

Moreover, novel authentication schemes designed to resist smudge attacks also increase the resilience towards thermal attacks since smudge attacks exploit a similar weakness in touch-based input. For example, SmudgeSafe [220] complicates smudge attacks against graphical passwords by randomly transforming the underlying image, causing the smudges to be different at every login attempt. Von Zezschwitz et al. [256] proposed three token-based graphical password schemes, two of which were significantly more secure against smudge attacks compared to patterns. The schemes rely on randomly positioned dragable objects. Hence thermal attacks are not expected to perform better than smudge attacks against these.

**Physical Protection Measures** While novel authentication schemes increase the resistance towards thermal attacks, increasing the security of current PIN and patterninput against thermal attacks is still an important aspect. Placing the hand on the display might remove all thermal traces on the screen as shown in Figure 5.3b. However, there are different procedures that decrease the success rate of thermal attacks without involving the user. For example, increasing the brightness of the display to the maximum for a few seconds heats up the display temperature and, thus, reduces the time thermal traces are visible, as depicted in Figure 5.3a. Similarly, running computationally heavy processes on the phone quickly heats the phone up, resulting in a similar effect as shown in Figure 5.3c.

### 4.4.7 Discussion and Summary

The results of our study and a review of prior work reveal that surfaces with specific properties can be used to detect on-surface interaction using thermal imaging. On this basis, in the previous sections, we presented the results from collecting and analyzing thermal traces of authentication processes, which we summarize and discuss grouped by the most important observations in the following.

We particularly focused on PINs and patternssince they are currently the most common knowledge-based authentication schemes [99, 254]. However, other authentication schemes may be vulnerable to thermal attacks as well. We expect attacks on graphical passwords that rely on cued-recall [19, 20, 220] to be similarly effective compared to the patternswe investigated. PINs are typically easy to uncover using observation attacks (De Luca et al. report 95% successful attack rate for PINs [58]). Our results indicate that PINs poorly resist thermal attacks as well, with overall success rates ranging from 78% to 100% when attacks are performed within the first 30 seconds after authentication (Figure 4.12). Although smudge attacks against PINs can uncover which digits were entered, hence significantly reducing the password space, thermal attacks can additionally uncover the order in which the digits were entered.

Without overlaps, patternsof maximum length are uncovered in 100% of the cases when thermal attacks are performed within 30 seconds after authentication (Figure 4.14). However, just adding one overlap significantly increases resistance to thermal attacks, as it influences the direction detection and the order of the performed patterns. Overlapping patterns did not have the same effect as duplicate PINs, as they also influence the detected direction and the order of the performed pattern. Hence we recommend including an overlap movement in patternsto increase resistance against thermal attacks.

In contrast to overlaps, knight moves do not distort the heat trace of pattern points but only the path at intersections. Hence, knight moves are as ineffective in making thermal attacks more difficult as they are against smudge attacks.

Moreover, unlike smudge attacks, *thermal attacks* do not require finding an optimal angle at which the traces are visible. Thermal attacks were shown to be tolerant to viewing angle/distance, as reported by Mowery et al. [174]. Mowery et al. evaluated different distances (30–70 cm) and did not observe changes in the detection. In our setup the camera was placed at 80 cm above the phone, hence we expect minimal to no influence of the distance on the results.

In contrast to observation attacks, thermal images are taken after authentication, hence the attack is less obvious to the victim and is not influenced by authentication speed. Additionally, the operation of thermal imaging allows seamless attacks, as it operates in a light invariant manner, where lighting conditions do not influence the capturing of thermal information [174].

Using a thermal camera with high temperature sensitivity and an automated computer vision approach to detect the traces, outperformed the results reported by prior related work [174]. Our approach unveils PINs/Patterns with high success after 30 seconds while previous work was successful up to only 3 seconds after authentication. While a higher sensitivity camera might have led to better results in manual analysis, we believe the main enhancement to come from the automated computer vision approach which allowed detection of heat traces despite being invisible to manual visual inspection.

In summary, we investigated the viability of thermal attacks on state-of-the-art touch screens and authentication schemes of mobile devices. To analyze the thermal images we implemented the ThermalAnalyzer, which was capable of uncovering 72%–100% of PINs in the first 30 seconds, and 100% of patterns that do not have overlaps. We additionally found that pattern overlaps significantly increase resistance to thermal attacks. Our work validates that thermal attacks are indeed a threat to mobile devices and should be considered by users and authentication scheme designers alike. We also furnish several solutions to protect from thermal attacks that are based on our results, previous work, and approaches to distort heat traces.

## 4.5 VID: Veins Patterns for User Identification

*"New technology is not good or evil in and of itself. It's all about how people choose to use it."*

**– David Wong –**

In the previous section, we presented an example of the potential threats thermal imaging may introduce. However, thermal imaging also opens novel opportunities for user identification based on the individual's vein patterns. In the following section, we present *VID*, contact-less vein-based identification.

Current biometric authentication mechanisms typically require the user to perform an explicit action. Fingerprint authentication, for example, requires a finger on a fingerprint sensor, and iris scanning requires looking into a camera. Another biometric authentication approach is the use of vein-patterns of the palm dorsal. Using vein patterns underneath the human skin has several advantages: Every person has a unique pattern of veins, which is stable from the age of ten and unique even for twins. Furthermore, as veins have a different temperature than the surrounding skin [52], thermal cameras can capture these patterns from a distance without interfering with the user's current task.

We investigate using veins on the back of the hand for contact-less and seamless user identification and authentication. In this section, we present the implemented vein-based authentication mechanism by combining thermal imaging and computer vision. Through a study we show that the approach achieves a low false-acceptance rate and a low false rejection rate. It is invariant to changes of the hand pose as well as to changes of the environment. Being accurate and invariant, vein-based authentication has the potential to be used to seamlessly authenticate users of desktop computers and tabletops. We demonstrate the potential of the approach through two prototypes that require no explicit action from the user.

## 4.5.1 Related Work

The uniqueness of the person's biometrics has been used to identify and authenticate users [50, 123]. Biometrics are classified into behavioral and physiological biometrics [282]. Behavioral biometrics relay on the behavioral cues to authenticate users, for instance their touch and keystroke patterns

Physiological biometrics is based on "something you are" and includes person's physiological information such as iris, face, voice, fingerprint, and hand geometry [231]. Various systems have been proposed which leverage these physiological information to identify users. A recent example is Bodyprint [110] that uses body parts like ear, finger, fist, and palm prints on a mobile phone touchscreen as an authentication scheme. Veins patterns are another biometric feature that has been proposed for identification and authentication. Since these pattern are unique for each users and stable for a lifetime, they are well suited as a biometric id. For example, finger veins can be captured under IR-lights [23] or vein pattern on the palm using VGA [259] or near IR imaging [83].

Researchers also explored using thermal (far IR) imagining to extract the veins on the palm [141, 260]. As veins have different temperature than the surrounding skin [52], they are visible to thermal cameras without any additional illumination sources. Vein triangulation and knuckle shapes are used to differentiate between users. Previous work, however, used a fixed hand pose recorded in controlled environments and even reported the inapplicability for outdoors environment. Being robust to pose variations and environment is, however, necessary to enable seamless and thereby usable identification and authentication.

## 4.5.2 Vein Identification

Identifying and authenticating users on the fly while interacting with interactive system is still a challenging task. We propose using a thermal camera that is already used for creating interactive systems [150, 210] to identify and authenticate users based on their vein pattern. Previous work shows that thermal imaging is well suited to detect vein pattern in a static setting by taking a thermal image of the user‘s hand while sitting in an controlled environment [260]. We focus on interactive systems that are controlled using different gestures and are deployed at different environment.

**Figure 4.17:** Veins extraction algorithm.

We propose a vein-based authentication approach, called VID. Using thermal imaging, the approach consists of a recognition pipeline that extracts vein patterns and authenticate users. For testing the approach we use an Optris PI450[23] contactless thermal camera. It has a $62° \times 48°$ field-of-view, $382 \times 288$ optical resolution, and temperature sensitivity of 40 mK. It should be noted that the temperature resolution is the most important criteria due to the small temperature differences between veins and the body.

## Recognition Pipeline

For extracting the hand veins, we are using the OpenCV library[24] for image processing and features extraction. We apply in total *six* steps to a thermal image stream to a identify a user as shown in Figure 4.17.

**1. Image extraction:** First, an image is extracted from the livestream of the thermal camera. This is currently done every 10 seconds but can be adopted to get faster identification.

**2. Noise filtering:** A $5 \times 5$ median filter is applied to smooth the image. The output is converted to grayscale and a 2D Gaussian filter is applied to further remove high frequency noise. Then, the image is normalized for better thresholding.

**3. Thresholding:** Extracting the veins (temperature contrast) using global thresholding did not yield in good results, as the vein-skin contrast changes over the hand. Hence, we use adaptive thresholding. The threshold value changes dynamically across the image with a kernel size $13 \times 13$. After this step, the result is a binary image of a user's hand.

[23] http://www.optris.com/thermal-imager-pi400

[24] OpenCV: http://opencv.org/

**4. Region Of Interest (ROI) extraction:** Our ROI covers the palm dorsal. To extract it, we detect the hand by finding the largest contour in the image. We compute the hand contour (depicted in green in Figure 4.17), convex hull, convexity defects, and the hand center (red circles in Figure 4.17). The ROI is computed to be the circle around the hand's center and a radius that is the distance between the center to farthest convexity defects. Hence, we can automatically compute the ROI without restriction to the angle of approach of the hand nor it's orientation.

**5. Veins Extraction:** After identifying the ROI, we segment the veins pattern and applied a morphology operation to reduce the thickness of the veins to a single pixel. Thus, the resulting image consists only of thin lines representing the veins. We use these images as features for our classification approach.

**6. Vein Classification:** For matching the extracted vein patterns, we calculated the Hausdorff distance [54] as a measure of similarity between two vein pattern. Based on a distance threshold, we determine the matching veins pattern using a nearest-neighbor approach.

### 4.5.3 Evaluation

To evaluate our VID approach, we conducted a user study in which we focused on two crucial aspects that are important for interactive systems. First, we investigated four different postures that are typically used when interacting (Figure 4.19) and, second, we looked into two different environmental settings, namely, indoors and outdoors. We chose these settings since we strive to have different setups that are especially challenging (e.g., for interaction in public spaces).

### Participants & Procedures

We recruited 29 participants (14 female, with an average age of 23 years, $SD = 3.6$) using our university's mailing lists. All participants were students studying different majors. We setup our VID system next to a table. The thermal camera was mounted on a tripod and faced toward the surface from a distance of 90 cm to capture the participant's hand (see Figure 5.1) similar to an interactive tabletop setup. The light and temperature conditions of the setup were constant during

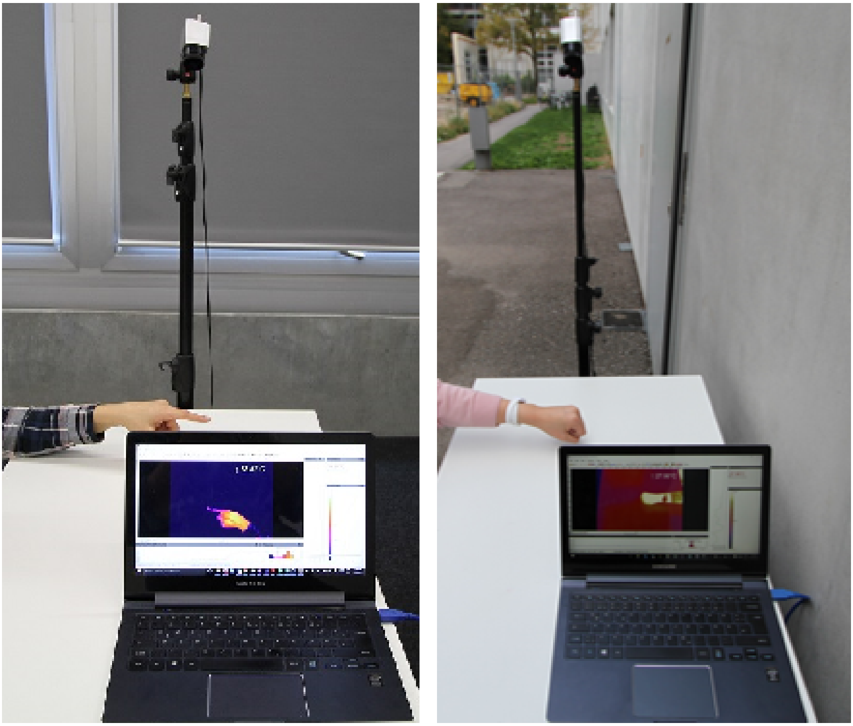


**Figure 4.18:** Study setup indoor and outdoor.

the whole study with ambient temperature of 23°C, 37°C for the indoors and outdoors setups respectively.

After welcoming the participants, we described the goal of the study and handed out consent forms as well as demographic questionnaires. We introduced the four different hand poses (Figure 4.19), but did not restrict them during the study with regards to neither the angle nor orientation of their hand. The study was conducted over two sessions in two days. During the first day, we captured thermal videos of the all participants' hands in 4 different poses indoors. The order of the performed poses were randomized per participant. We recorded a thermal video of 30 seconds length for each pose. We deliberately chose to record videos rather than doing live authentication because of the reproducibility of our evaluation. The study took approximately 20 minutes per participant. The second day, 15 of the participants where re-invited to repeat the same poses using the

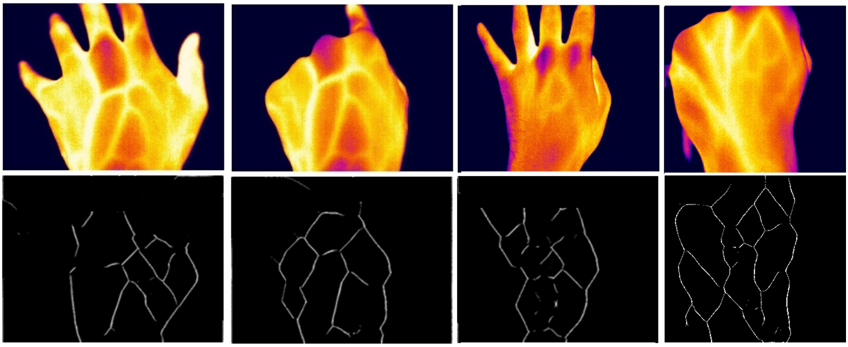


**Figure 4.19:** Thermal view and extracted veins pattern for the four hand poses.

same setup but outdoors, aiming to evaluate the environmental influence on our system. The procedure was identical to the first day.

### 4.5.4 Results

We replayed the recorded videos to assess our proposed approach. Our analysis included the evaluation of the influence of (1) different hand poses, and (2) environmental changes on the authentication accuracy.

**Analyzing Hand Pose** In the first step, we analyzed the influence of the different hand poses. We trained our system using fourfold cross validation with one specific hand pose in each fold. For each user, we had 9 recordings of 3 hand poses in the training set and 3 recordings of one hand pose in the test set. Next, we used this data to determine the False Rejection Rate (FRR).

In the second step, we calculated the False Acceptance Rate (FAR). We followed the same approach as above but left additionally one participant completely out of the training set. Thus, this unknown potential user (attacker) tried to authenticate with all poses (i.e., with 12 recordings). We repeated this procedure for all participants.

Based on the computed FAR and FRR, we calculated the Equal Error Rate (EER) which is typically used to quantify the performance of an authentication system [123]. The results of the FRR and FAR against different threshold values are depicted in Figure 4.20. We achieved an EER of 17%. Over total 29 users,

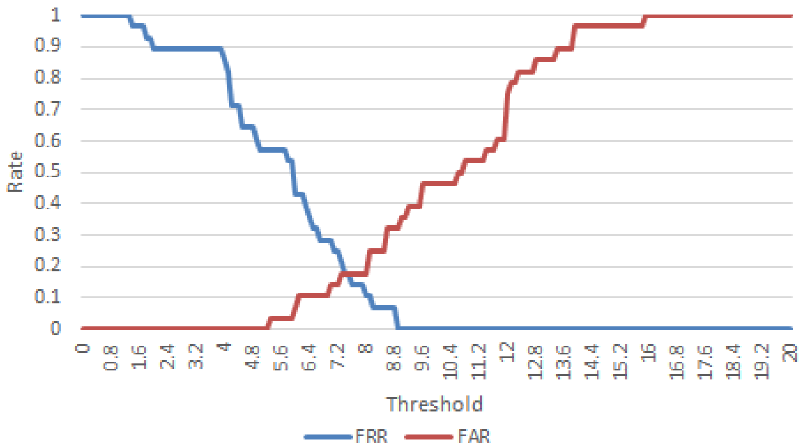



**Figure 4.20:** False acceptance and the false rejection rates.

with the acceptance threshold of 9, this authentication scheme can achieve an identification accuracy up to 100%.

**Analyzing Environmental and Temporal Influences** We further investigated the influence of the environmental and temporal changes. We used the recordings from the 15 participants that have been re-invited on the second day of the study. We trained our system using cross validation with one setup and session in each fold. In contrast, we used the recordings from the first day for training the users and the recordings from the 2nd day for querying. The FAR and FRR are computed using the same procedure described earlier. We achieved an EER of 7%. With the acceptance threshold equal to 9, we achieved an identification accuracy of 100%.

## Application Scenarios

We envision two different application scenarios in which our approach can implicitly identify and/or authenticate users.

**Tabletop Multiuser Identification** . Realizing interactive tabletops using thermal imaging yields several advantages such as the possibility to detect the amount of pressure applied on the surface or the traces made by the user [150]. In addition to that, our approach might enrich such a system by providing user

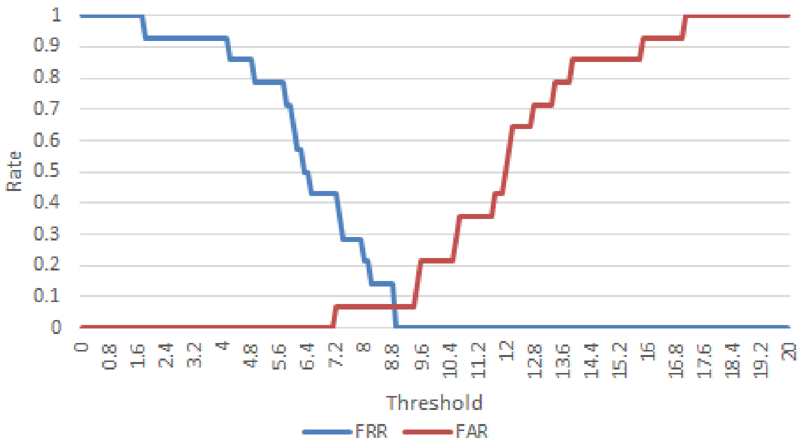



**Figure 4.21:** False acceptance and the false rejection rates.

identification so that multiple users are able to interact at the same time and the system is capable of differentiating them (Figure 4.22). Further, it is possible to easily differentiate between input from the left and right hand since their veins pattern differs for the same user. This could be used to extend the gesture space, where the hand used to perform the gesture specify the action performed. For instance, if the user authenticates with the left hand the system logs him/her in to the public account and if the user uses the right hand it logs into the private account.

**Laptop Authentication.** Since thermal cameras are becoming smaller[25], we envision that the integration of this type of cameras into the display of a laptop will become a reality. Having a thermal camera that faces down the keyboard as shown in Figure 4.22, our system is able to track who is using the laptop and can reject the access for unknown users to specific information. At the same time, it preserves the privacy of the user since thermal imaging is perceived as a temperature sensor, which is more accepted compared to RGB ones.

[25] http://www.flir.com/

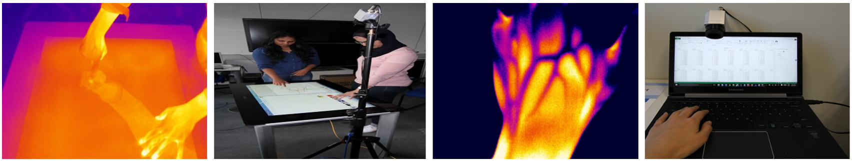


**Figure 4.22:** Potential Use-cases.

### 4.5.5 Discussion and Summary

The results of the conducted study, highlights the potential of using veins patterns as bases of implicit biometric authentication for interactive systems. We show that it's possible to distinguish users regardless of their hand pose and the ambient temperature of the surroundings with an EER that is on par with recently proposed approaches [110]. Hence, a convenient, usable, continuous authentication scheme can be considered. It doesn't interrupt the user's task, as it seamlessly capture the veins on the palm dorsal. It also address hygiene issues, as the veins are captured in a contactless manner. Relying on the vein pattern underneath the skin makes it tolerant to skin conditions such as grease and wet hands. However, user acceptance should be investigated, as user have concerns about the storage of their physiological biometrics [198].

Relying on biometric features, as replacement of passwords and token, enhances the convenience of authentication. Users usually interact with devices like laptops and interactive tabletop without occluding the back of their hand. Thus, deploying thermal camera, which currently become affordable and small, to capture their veins pattern allows user identification and authentication. It is seamless, contactless, and continuous authentication scheme. Based on the approach proposed the user identification can be up to 100% precision with different hand poses both in indoors and outdoors settings.

# Chapter 5

# Amplified Perception of Cognitive Load

*This chapter is based on the following publications:*

- Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere. Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technology*, 1(3):33:1–33:20, Sept. 2017
- Y. Abdelrahman, A. Khan, J. Newn, E. Velloso, S. Safwat, J. Bailey, A. Bulling, F. Vetere, and A. Schmidt. Classifying attention types with thermal imaging and eye tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technology*, Sept. 2019

Humans exhibit great abilities and social intelligence in their ability to discern emotions and internal states. According to Picard et. al, emotional intelligence consists of "the ability to recognize, express, and have emotions, coupled with the ability to regulate these emotions, harness them for constructive purposes, and skilfully handle the emotions of others" [196].

However, the limited awareness of our cognitive load levels can have a great effect on our health, rational thinking, intelligence, and ability to react and adapt to what is important. Currently, we as well as digital systems are largely blind to users' cognitive states. Systems that can assist in this respect show great potential for augmenting cognition and for creating novel user experiences. However, most approaches for sensing cognitive states, and cognitive load specifically, involve obtrusive technologies such as physiological sensors attached to users' bodies.

Building systems that extend our cognitive abilities, augment our intellect [72], work in symbiosis with humans [155], and provide ubiquitous access to information [264] has been a core theme in human-centered computing since its inception. These aspirations have carried on through multiple research programs, including Affective Computing [196], Physiological Computing [75], and more recently Symbiotic Interaction [122] and Human Amplification [215]. These cognition-aware systems aim to sense users' internal states and to adapt their interface and behaviour accordingly. Such systems offer opportunities to tailor educational activities in online learning environments, to dynamically optimize work-flows for knowledge, to improve performance for assembly line workers [85], and to focus users' attention in critical systems. A crucial step in building cognition-aware systems is capturing different aspects of users' mental states, such as their cognitive load, loci of attention, and affect. Despite over 50 years of work in the area, how to sense cognitive load in a robust, accurate, timely, and unobtrusive way is still an open challenge.

Cognitive load has been measured traditionally in two ways: (1) by subjective self-reporting and (2) by observing user performance in a task or in a set of parallel tasks. The NASA TLX is a common example of the first category, where participants are asked to report their own load with regard to 6 different categories. Another example where study participants are asked to report their own estimates can be found in Sweller et al. [239]. The drawback of these approaches is that the answers are highly subjective. Furthermore, the self-reporting itself adds to the cognitive load. Measuring cognitive load through the performance in the task itself or in a secondary task (e.g. Lane Change Task for Automotive user interface, ISO 26022) only provides a rough estimate and is typically only suitable to laboratory studies and not for creating cognition-aware real-time systems. For interactive systems to be able to adapt their behavior accordingly, cognitive load information must be captured continuously and automatically. Introspection is often not sufficient. Physiological sensors such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and galvanic skin response (GSR) sensors show potential as possible solutions to this problem, but are limited in their application in ubiquitous computing environments since they require users to wear obtrusive additional hardware (e.g. electrodes on their skin).

By adding sensing capabilities to the environment rather than burdening the user, thermal imaging is a strong candidate for the task of measuring cognitive load. Thermal cameras are both unobtrusive and able to capture information from multiple users at a distance and at the same time. Previous research has shown that thermal patterns reveal different aspects of our internal states, including affect [117, 235], stress [203], and deception [206]. Further, advances in miniaturization and mass production have continuously brought down the prices of these devices. With consumer-grade cameras readily available in the market for a few hundred dollars, measuring cognitive load at a larger scale becomes feasible.

In this chapter we present our exploration to assess the effect of different levels of cognitive load on facial temperature. Two study probes target cognitive load levels estimation. In Section 5.1, we explore a novel method for estimating cognitive load based on users' facial temperature patterns using a commercial thermal camera. We then extend our exploration to investigate the effect of different attention types on the cognitive load, hence, acquiring additional cue to classify attention types as presented in Section 5.2.

The research question we address in the presented chapter is:

- **RQ4**: How can thermal imaging be used to amplify perception of cognitive load?

## 5.1 CognitiveHeat: Estimating Cognitive Load Level

*This section is based on the following publications:*

- Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere. Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technology*, 1(3):33:1–33:20, Sept. 2017

Our cognitive and affective states strongly influence how our blood flows through our bodies. When we are scared, blood flows to our legs in reaction to the *fight or flight* response; when we are embarrassed, blood flows to our face, making us blush. Because blood carries heat, as it flows through our bodies it changes the temperature distribution in our skin [229], underlying tissues, and vessels [223, 229]. Therefore, monitoring changes in this distribution can give us an insight into the changes in cognitive load or arousal that caused them.

Previous works have suggested several points in the body to measure this temperature fluctuation, such as the nose, the cheeks, the areas around the eyes (periorbital and supraorbital), the jaw, the neck, the hands (fingers and palms), the lips, and the mouth [117]. Thus the face is particularly promising for this task for several reasons. First, it is often exposed, making it easy to observe with a thermal camera. Second, it features a thin layer of tissue, making temperature changes more pronounced. Therefore, in this work we explore how facial temperature fluctuations can give us an insight into changes in cognitive load. We focus on two of the points of interest suggested in the literature—the forehead and the nose—as these can be monitored even if the user is wearing glasses. In summary, in this work we focus on the following:

1. Can we distinguish different cognitive load levels using a relatively low-cost, commercially available thermal camera? More specifically, do the changes in facial temperature correlate with the level of difficulty of the task?
2. If so, how long after the increase in cognitive load can we detect the corresponding temperature fluctuation? In other words, what is the latency of our method as a cognitive load sensor?

## 5.1.1 Related Work

Awareness of cognitive load and processes is an ongoing challenge for HCI research. Our work builds on two strands of prior work: (1) cognitive load estimation and (2) thermal imaging.

### Cognitive Load Inference

A technology with the capability of sensing and inferring cognitive load has the ability to provide a "window into our mind" that can further be used to adapt system behavior accordingly [245, 92, 91]. However, capturing cognitive state information is a challenging task—cognitive processes are largely invisible from the outside of the users' brain and introspection often fails to reason about them in a unbiased and objective way.

Monitoring users with the help of sensors can give us clues about different cognitive states. Though certain physiological sensors can be highly specialized, expensive, and therefore only applicable under lab conditions, advances in sensor technology have led to inexpensive solutions that can be easily integrated into personal devices.

Different approaches have been introduced to infer cognitive states ranging from using facial expressions [261], eye movements [98] and pupil size [29, 194], skin conductance [106, 146, 226], brain signals [53, 227, 228], Electrodermal activity (EDA) [188, 189, 93], heart and respiration rate [146].

**EEG:** Haak et al. [98] reported that the blinking rate is directly proportional to cognitive load. They extracted the blinking rate from Electroencephalography (EEG) signals by isolating events from the signals. Hosni et al. [112] used EEG signals for task classification. Shirazi et al. [228], classified reading and relaxing tasks based on EEG signals retrieved from a single electrode BCI. Petersen et al. [193] used the EPOC to distinguish emotional responses when viewing different content.

**Galvanic Skin Response (GSR):** Analysis of GSR data from user experiments has shown that GSR across users increases along with cognitive load [71]. Yu et al. explored the applicability of using GSR as an indicator of cognitive load [226]. Elise et al. used heart rate, respiration rate and GSR as an indicator [146].

**Respiration Rate:** Parnandi et al.considered real-time adaptive biofeedback games [188, 189] . They monitored players' EDA to infer their arousal states [188]. Additionally, they used biofeedback sensors (respiration rate sensor and adaptive games) to manipulate their behavior [189].

**Combined Sensors:** Wang et al. [261, 71] explored how to build an adaptive system that helps workers who use computer heavily on a daily basis by extracting a user's features, such as face pose, eye blinking, yawn frequency and eye gaze from a recorded video, in order to monitor the users state. Healey et al. [104] and Schneegass et al. [219] used physiological monitoring for driver stress indication.

A major limitation of both voice and facial-based approaches is that users can be quite skilled at manipulating the parameters being sensed by the system. On the other hand, physiological metrics, such as heart rate, GSR, Blood Volume Pressure (BVP), and Electromyography (EMG) have the advantage that they are primarily under the control of the Autonomic Nervous Systems (ANS) and are therefore less susceptible to conscious manipulation. However, a major limitation of current physiological approaches is the need for sensors to be in direct contact with the user, or to be implanted. As a result, such sensors are impractical for most routine user environments. For instance, the long setup time and contact requirements of BCIs, or the drift over time [153] and fluctuations due to arm movements in GSR.

As promising tools within the HCI domain, thermal cameras show high potential for estimating cognitive load. They overcome the limitation of using contact sensors utilized in previous research and are more robust than other contactless approaches, since the temperature signature is more resistant to conscious manipulation [117]. We investigate the use of thermal imaging to estimate cognitive load, exploiting the fact that cognitive load influences the skin temperature (which is directly related to the conduction of heat from the blood to the facial skin [191]) as a reflection of the activation of the ANS [73]. We therefore aim to leverage the correlation between cardiovascular physiology and mental state, where they are capable of reliably differentiating between levels of cognitive load [229].

Thermal cameras can provide information about the observed body's temperature, which can be used to infer the physiological [117] and cognitive state of users in an unobtrusive manner by, for example, evaluating their stress levels [203]. The reason why this is possible is because our skin temperature is modulated by ANS activity. ANS controls the organs of our body, such as the heart, stomach, and intestines. It is responsible for activating glands and organs for defending the body from threats. Its activation might be accompanied by many bodily reactions, such as an increase in heart rate, rapid blood flow to the muscles, activation of sweat glands, and increase in the respiration rate. These physiological changes can be measured objectively by using sensors [143, 222].

Temperature changes on the forehead have been shown to be linked to changes in brain temperature [89, 169]. There is a direct relationship between workload and facial temperature based on the involvement of the autonomic nervous system (ANS) [115]: increased brain activity causes a surge in blood supply. Hence, higher workloads lead to blood flowing from the adjacent facial areas to the brain causing the facial temperature to vary. Zajonc et al. [288] showed different facial areas to be effective temperature indicators, namely the tip of the nose, above the eyes, and at the center of the forehead.

Previous works have explored the usage of thermal imaging to observe users' mental states. We particularly build upon previous work that assessed stress based on the variations in forehead and nose temperature [73, 192, 224]. Emotions like stress [203], fear [170], startling [223], empathy [165], anxiety [290], and guilt [116] could be detected by monitoring facial temperature changes. Ioannou et al. summarized these states and how they correlate to the facial temperature in terms of region of interest and direction of temperature change (i.e. increase or decrease in the temperature) [116].

Compared to previously established sensors, the great advantage of thermal imaging is its contactless and non-invasive operation. The contact-free recording of facial temperature with an thermal camera allows us to isolate unsystematic data variation (e.g. users' bias due to their awareness of being monitored, the movement of the sensor or the stressful attachment of the sensor on the users' body). Additionally, instrumenting the environment is more user-friendly and allows the tracking of multiple users. Most of the research so far has used MWIR thermal cameras. For instance, *StressCam* [203] used the Indigo Phoenix thermal camera costing over $20,000$ USD. Jenkins and Brown [124], utilized the supraorbital region to identify cognitive state, yet they used a non-commercial thermal camera. However, because FIR thermal cameras are commercially available and relatively affordable, they present a compelling opportunity for expanding the reach of these applications. For instance, it is now possible to buy smartphones with built in thermal cameras [26]. These cameras are becoming increasingly smaller, with sizes as small as a 20mm [27], yet maintaining high thermal sensitivity around $0.05^{\circ}$ degrees. This enables thermal cameras to be used in a diverse set of applications, by enhancing existing application scenarios and exploring new ones. Previous research has shown that thermal imaging in the

[26] http://www.catphones.com/en-gb/phones/s60-smartphone

[27] http://www.flir.com/cores/lepton/

NIR and MWIR bands can be used to reveal different cognitive states. Stemberger et al. [235] explored the use of FIR to estimate cognitive load levels, but they used a wearable tracking headset to identify the region of interest, and a neural network to build a user dependent classification to three-levels of workload based on six region of interest. Or and Duffy [186], and Kang et al. [128, 129], used the variation in the nose temperature as an indicator of cognitive workload. However, they didn't report on how different levels of workload influence the temperature change. Additionally, their findings are confounded by facial temperature stress indicators.

In this work, we explore thermal imaging operating in the FIR band. We leverage the advances in miniaturization and reduction in the prices of these devices, to explore the feasibility to not only detect the cognitive load, but also to estimate four levels of cognitive load, while maintaining the unobtrusive operation manner of thermal imaging. We aim to investigate the possibility of using thermal imaging as a user-independent cognitive state detector. Additionally, we also explore different metrics to avoid any possible overlap between other states e.g. stress.

In summary, the aim is to address two major shortcomings of previous work concerned with estimating cognitive load: The obtrusive and contact nature of traditional physiological sensors and the limitation in detecting different cognitive load levels in a user-independent manner.

### 5.1.2 Thermal Imaging for Cognitive Load Estimation

We built a system that monitors users' facial temperature (see Section 5.1.3) and observed how it changed as users performed two tasks with four levels of difficulty each (see Section 5.1.3). We hypothesized that the higher task difficulty will result in a greater temperature difference. Further we conducted a second study and measured how long before the temperature started to change after the task started or ended, and how long it took for it to reach its maximum level (see Section 5.1.4). We present our results regarding the applicability of thermal cameras as a cognitive load sensor.

### 5.1.3 Cognitive load and Facial Temperature

To test our hypothesis of the ability of thermal cameras to elicit cognitive states and classify tasks based on face temperature variation, we conducted a user study in which we recorded the temperature of participants' nose and forehead in three activity states:

1. Relaxing as the baseline.
2. Reading four different types of text.
3. The Stroop test [236] with four levels of difficulty.

## Design

We applied a repeated-measures design, where all participants were exposed to all three task conditions. We studied the effect of the tasks on the facial thermal print. For the baseline we asked the participants to relax. For the reading task we provided four types of different content types: 1) a comic, 2) an easy blog article, 3) a scientific article and 4) a literary piece. We chose these content types because of their presumed differences in cognitive demand. Additionally, we computed the readability index [28] for each text, which indicates the text difficulty: the higher the value the more difficult the text is to read. The text found in the comic, easy blog, science article and literary piece reported 26.6, 52.9, 68.2 and 77.9 respectively.

The Stroop test is a classic Psychology task for evaluating executive functions [236]. During the test, users are asked to name the color of the font in which different words are written. The difficulty of the task lies in the fact that the words displayed represent a different color to the one in which it is colored. For example, the word 'red' would appear colored in blue, and the participant had to say 'blue'. In our study we also introduced four levels of difficulty in the task by adding four levels of increasing time pressure: the higher the level, the less time users had to respond. For varying the difficulty of the Stroop test, we considered four levels of difficulty provided by the app *Magic Colors*[29]. To overcome the effect of the repeated-measures experimental design, namely order effect, the order of the tasks was counter-balanced using a Latin Square.

[28] https://www.psychometrica.de/lix.html

[29] https://play.google.com/store/apps/details?id=com.accountmaster.in.MagicColors

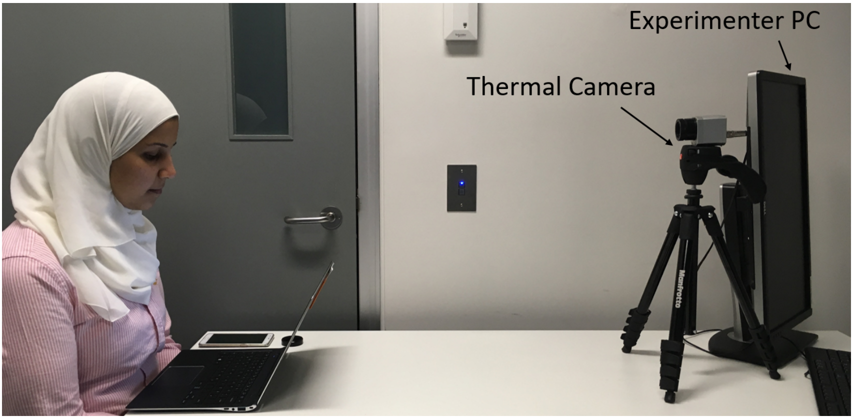



**Figure 5.1:** Study setup.

## Apparatus

Our setup consisted of a 13.3" laptop in front of a thermal camera (Optris PI160[30]) mounted on a tripod. The optical resolution of our camera was 160×120 pixels and its frame rate was 120 Hz. It is able to measure temperatures between -20°C and 900°C, and operates with a thermal sensitivity of 0.08°C represented by the noise equivalent temperature difference (NETD)[31]. The wavelengths captured by the camera are in the spectral range between 7.5$\mu$m and 13$\mu$m. The lens we use provides a 23°× 17° field of view. The thermal camera uses USB as power source as well as to transfer data. It provides temperature information in the form of 16-bit color values encoding the temperature information. The participants were asked to look to the front facing the thermal camera placed at 1m from the participants and the screen as shown in figure 5.1.

## Implementation

We built a system that recognizes and analyzes the user's facial temperature. Our application receives the data from the thermal camera, and extracts the temperature of the forehead and nose as follows:

[30] `http://www.optris.com/thermal-imager-pi160`

[31] NETD refers to the electronic noise that is interpreted as a temperature difference of an object

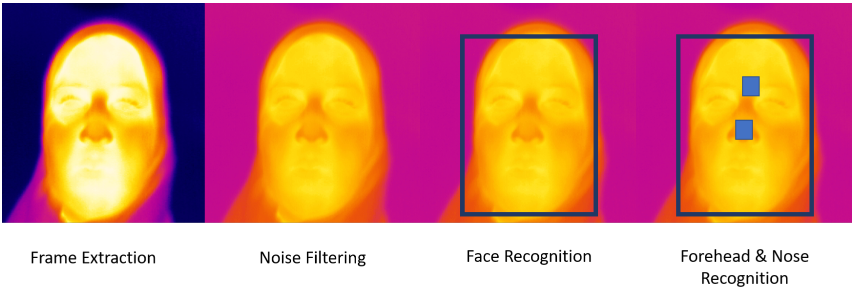



**Figure 5.2:** Nose and Forehead ROI extraction.

1. **Frame extraction:** We sample each frame of the camera feed at 120fps, based on the cameras frequency.

2. **Noise filtering:** We apply a $5 \times 5$ median filter to smooth the image. We convert the output to gray-scale and apply a 2D Gaussian filter to further remove high frequency noise as performed by Shirazi et al. [210].

3. **Face Recognition:** We detect faces in the frame using the Viola-Jones classifier [248] built into OpenCV.

4. **ROI Identification:** We identify the nose tip and forehead as the ROI. These ROI are computed relative to the face coordinates extracted as shown below. We used a simple ROI identification approach to maintain a fast frame rate for the algorithm.

```
xForehead = xFace + (4 * face.Width / 7);
yForehead = yFace + (face.Height / 6);

xNasal = xFace + (4 * face.Width / 9);
yNasal = yFace + (face.Height / 2);
```

5. **Temperature Recording:** We record the average temperature of the 5x5 window CSV file to represent the temperature of the nasal tip and forehead, as well as the difference in temperature between the two.

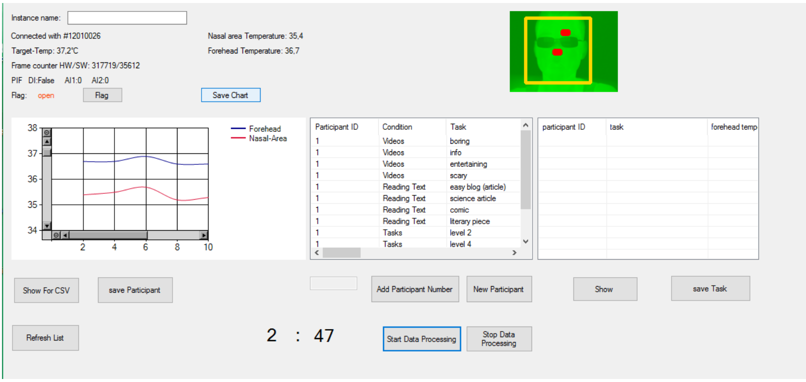



**Figure 5.3:** The application interface used during the study.

## Participants and Procedure

We recruited 12 participants (7 females) with an average age of 28.3 years ($SD = 4.6$) using university mailing lists. None of the participants had any previous experience with thermal cameras. After arriving in the lab, participants signed a consent form and received an explanation of the purpose of the study. Next, we asked participants to perform the set of reading tasks, video watching and Stroop tests, each for 12 minutes (3 mins $\times$ 4 levels). The order of the tasks was counter-balanced using Latin-square. The study took approximately 60 minutes. During the entire experiment, we recorded the temperature of the participant's face, extracting the forehead and nasal temperatures.

We recorded the whole study using an RGB video camera. The experiment was conducted in a maintained room temperature of 24°C. Participants were rewarded with 10 EUR. All the data was visualized by the experimenter in real-time in an accompanying application developed in C# (see Figure 5.3).

| Reading Task | Nose | Forehead | Forehead-Nose Temperature |
|---|---|---|---|
| Comic | -0.69 | 0.23 | 0.92 |
| Blog | -0.92 | 0.33 | 1.25 |
| Article | -1.11 | 0.37 | 1.48 |
| Literature | -1.49 | 0.44 | 1.93 |

**Table 5.1:** Mean temperature change in the Reading tasks.

## Results

We analyzed the effect of the task difficulty on the recorded facial temperature. We used three metrics as our dependent variables:

1. Decrease in nose temperature.
2. Increase in forehead temperature.
3. Difference between nose and forehead temperature.

We defined the temperature change as the difference between the mean temperature during the baseline recording and the mean temperature in the final minute of the task.

## Effect of Reading task on ROI Temperature

**Nose Temperature** We tested the effect of the CONTENT DIFFICULTY on the NOSE TEMPERATURE with a one-way ANOVA. Mauchly's test showed a violation of sphericity against CONTENT DIFFICULTY (0.07, p<.05), so we report Greenhouse-Geisser-corrected ($GGe = 0.45$) values. We found a large significant effect of CONTENT DIFFICULTY on the NOSE TEMPERATURE ($F_{1.35,14.9} = 14.0, p < .0001, ges = 0.29$). Bonferroni-corrected pos-hoc tests found a statistically significant difference between all content types ($p < .05$), except between the blog and the science article, and between the science article and the literary piece at $p < .05$. The mean decrease in temperature between levels was of .27 degrees Celsius.

**Forehead Temperature** We tested the effect of the reading CONTENT DIFFICULTY (4 levels) on the FOREHEAD TEMPERATURE (difference to the baseline) with a one-way ANOVA. Mauchly's test showed a violation of sphericity

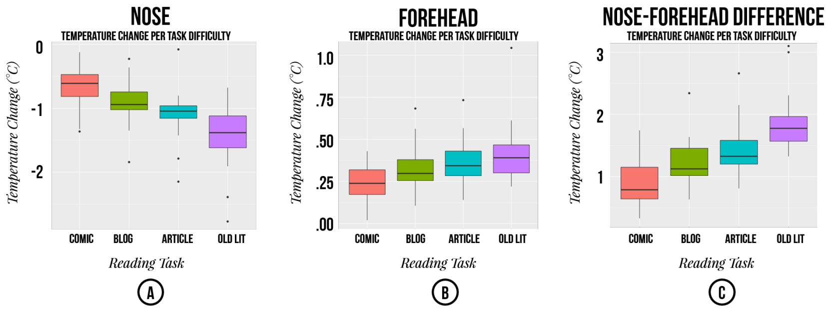



**Figure 5.4:** Temperature change between the baseline and the Reading tasks.

against difficulty $(0.01, p < .05)$, so we report Greenhouse-Geisser-corrected $(GGe = 0.37)$ values.

We found a significant large effect of CONTENT DIFFICULTY on the FOREHEAD TEMPERATURE $(F_{1.12,12.33} = 19.78, p < .001, ges = 0.16)$. Bonferroni-corrected pos-hoc tests found a statistically significant difference between all content types $(p < .05)$, except between the science article and the literary piece. The mean increase in temperature between levels of difficulty was .07 degrees Celsius.

**Forehead-Nose Temperature Difference** We tested the effect on the difference between forehead and nose temperature (difference to the baseline) with a one-way ANOVA. Mauchly's test showed a violation of sphericity against CONTENT DIFFICULTY $(0.09, p < .05)$, so we report Greenhouse-Geisser-corrected $(GGe = 0.44)$ values. We found a large significant effect of CONTENT DIFFICULTY on the FOREHEAD-NOSE DIFFERENCE $(F_{1.32,14.54} = 23.26, p < .0001, ges = 0.38)$. Bonferroni-corrected post-hoc tests found significant differences between all levels of difficulty. The mean increase in temperature difference between the forehead and the nose between levels was of $.34(\pm 0.12)$ degrees Celsius.

In summary, our reading tasks exhibited a significant increase in the forehead temperature and decrease in the nasal temperature. We found a significant difference between all contents for the increase in the forehead-nose temperature difference, and a larger effect size of the task difficulty on this metric. The difference between levels of difficulty in the order of .34 degrees Celsius.

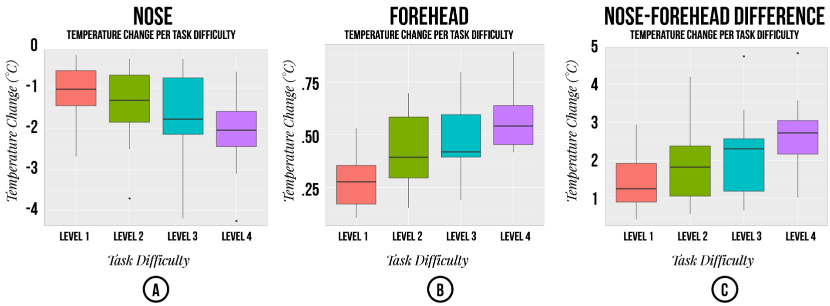



**Figure 5.5:** Temperature change between the baseline and the Stroop tasks.

## Effect of Stroop Task Levels on ROI Temperature

**Nose Temperature** We then tested the effect of the TASK DIFFICULTY (4 levels) on NOSE TEMPERATURE (difference to the baseline) with a one-way ANOVA. Mauchly's test showed a violation of sphericity against TASK DIFFICULTY ($0.08, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.68$) values. We found a large significant effect of task difficulty on NOSE TEMPERATURE ($F_{2.05,22.57} = 29.1, p < .0001, ges = 0.14$). Bonferroni-corrected post-hoc tests found significant differences between all levels of difficulty. For each increase in the level of difficulty we found an decrease of 0.33 ($\pm 0.12$) degrees celsius in nose temperature as estimated by a linear regression model.

**Forehead Temperature** We tested the effect of the TASK DIFFICULTY (4 levels) on the FOREHEAD TEMPERATURE (difference to the baseline) with a one-way ANOVA. A Mauchly's test showed a violation of sphericity against Difficulty ($0.06, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.58$) values. We found a large significant effect of TASK DIFFICULTY on FOREHEAD TEMPERATURE ($F_{1.73,19.01} = 14.99, p < .001, ges = 0.31$). However, Bonferroni-corrected pos-hoc tests did not find a significant difference between levels 1 and 2, and between levels 3 and 4. For each increase in the level of difficulty we found an increase of 0.09 ($\pm 0.02$) degrees Celsius in the forehead temperature as estimated by a linear regression model.

**Forehead-Nose Temperature Difference** We tested the effect of the TASK DIFFICULTY (4 levels) on the difference between forehead and the nose temperature (difference to the baseline) with a one-way ANOVA. Mauchly's test showed a violation of sphericity against TASK DIFFICULTY ($0.18, p < .05$),

| Stroop Level | Nose | Forehead | Forehead-Nose Temperature |
|---|---|---|---|
| Level 1 | -1.11 | 0.29 | 1.40 |
| Level 2 | -1.45 | 0.42 | 1.87 |
| Level 3 | -1.68 | 0.48 | 2.16 |
| Level 4 | -2.12 | 0.57 | 2.69 |

**Table 5.2:** Mean temperature change in the Stroop tasks.

so we report Greenhouse-Geisser-corrected ($GGe = 0.68$) values. We found a large significant effect of TASK DIFFICULTY on FOREHEAD-NOSE DIFFERENCE ($F_{2.03,22.36} = 37.97, p < .0001, ges = 0.20$).Bonferroni-corrected post-hoc tests found significant differences between all levels of difficulty. For each increase in the level of difficulty we found a difference of 0.42 ($\pm$0.13) degrees Celsius in the difference between forehead and nose temperatures as estimated by a linear regression model.

In summary, we found statistically significant effects of task difficulty on temperature measures on (1) forehead, (2) nose, and (3) the difference between the forehead and the nose. The largest effect was found in the difference between the forehead and the nose.

Informed by previous work, we hypothesized that an increase in the task difficulty would lead to a change in the participants' facial temperature patterns. Because forehead and nose are two of the most visible points on users' faces and are two points recommended deemed feasible for temperature measurement by previous work [117], we tested the effects of different tasks and their difficulties on temperature changes in these points.

We elicited increases in cognitive load both through an abstract task and through a naturalistic task. Our abstract task consisted of a variant of the classic Stroop test, in which we increased the task difficulty by introducing a time pressure. In this task, we found that an increase in the task's difficulty lead to a change both in forehead and nose temperature. The corresponding changes were related—an increase in task difficulty lead to an increase in forehead temperature and a decrease in nose temperature. We therefore combined both metrics by calculating the difference in temperature changes between the two. This proved to be the most robust metric, with a statistically significant average increase of 0.42 ($\pm$.13) degrees Celsius between each difficulty level.

We confirmed the validity of this finding in a naturalistic scenario, consisting of reading four pieces of text with varying levels of difficulty as measured by a readability scale—a comic book, a blog post, a scientific article, and a snippet of Old German literature. Again, we found a significant effect of the task difficulty for all metrics, in the same directions as in the Stroop task. Though this difference was not significant for all pairs of tasks in the forehead and nose temperatures alone, they were significant for all pairs when combining the two by subtracting the latter from the former. We found an average increase in the temperature difference between the forehead and the nose of .34 degrees Celsius for between each level of difficulty.

The forehead temperature increases are correlated with metabolic increases in this ROI. This is presumed to be due to the influence of muscle activation of the forehead muscle group [185, 229]. In parallel, the vessels in the nose region experience vasoconstriction (tightening in the blood vessels) as response to increased cognitive load [124, 229], reflecting a decrease in nose temperature.

### 5.1.4 Temporal Latency of Facial Temperature Change

Our first study validated the suitability of using the temperature differences between forehead and nose as a metric for cognitive load sensing. We found that the temperature changes are large enough for some of the cheapest thermal cameras in the market to capture. In the first study, we were interested in the *magnitude* of the temperature changes and therefore, we were only concerned with the average temperature at the end of the tasks. However, in a realistic scenario, we would be interested in pinpointing specific times in which changes in the facial thermal pattern could be detected. This would allow us to build cognition-aware systems that detect user state changes in real-time. For this purpose, it is crucial to understand the temporal response of these changes, which was the focus of our second study.

Other physiological sensors like GSR exhibit response times around three seconds [142]. The response latency achieved in previous works with thermal imaging include 10secs [142] on monkeys subjects and 3.8secs using functional thermal imaging [170]. The high latency found in previous works are not ideal for real-time applications. In our work, we wanted to investigate the latency, thereby investigating whether the current state of commercially available thermal imaging is appropriate for measuring cognitive load levels in real-time and in real world cognition-aware applications. To the best of our knowledge, no work has been done in evaluating the temporal latency of temperature changes


**Figure 5.6:** Example for one study sequence.

using commercial thermal cameras which operate in the far infrared spectrum, particularly considering different stimuli duration. We evaluated the response time of the temperature change while considering different task duration ranging from 5 to 60 seconds.

## Design

For this, study we applied a repeated-measures design, where all participants were exposed to all conditions. We studied the effect of the duration of the task on the latency of the temperature change. We chose the Stroop test as the task/stimulus, with task duration of 5, 15, 30, 45 and 60 seconds. Each duration value was repeated three times.

## Apparatus

The general setup was the same as in the first study, except that we used a more precise thermal camera with higher thermal sensitivity: the Optris PI450[32] with an optical resolution of 382×288 pixels and a frame rate of 80 Hz. It measures temperatures between -20°C and 900°C and operates with a thermal sensitivity of 0.04°C represented by the noise equivalent temperature difference (NETD). The lens we use provides a 38°× 29° field of view.

## Participants and Procedure

We recruited 24 participants (13 females) with an average age of 30.8 years ($SD =$ 9.6). The participants were two groups: native Egyptians and Canadians. None of the participants had any previous experience with thermal cameras. Participants first signed a consent form and the purpose of the study was explained to them. Next, we asked them to relax for 10 minutes to ensure no other factors influencing

[32] `http://www.optris.com/thermal-imager-pi400`

| Stroop Level | Nose | Forehead | Forehead-Nose Temperature |
|---|---|---|---|
| Onset | 0.7(±0.2) | 1.2(±0.3) | 0.7(±0.2) |
| Saturation | 3.1(±1.2) | 2.3(±0.9) | 2.3(±1.2) |
| Offset | 1.1(±0.5) | 1.6(±0.9) | 1.1(±0.5) |

**Table 5.3:** Summary for the onset, saturation and offset in seconds.

the facial temperature for instance rushing into the study room. We recorded baseline temperature measures while participants were relaxing. They were then introduced to the Stroop task with different exposure duration, each with three minutes break between them. The order of the duration was counter-balanced using a Latin square. The study took approximately 60 minutes. During the entire experiment we recorded the temperature of the participant's forehead and nasal area. The experiment was conducted in a maintained room temperature of 26°C.

## Results

We analyzed the effect of the different durations of the stimuli/tasks on the latency of facial temperature variations. As in study I, We used the same three metrics as our dependent variables: nose temperature, forehead temperature and differential temperature. We investigated the following:

1. Temperature change ***onset***, refers to the time taken to first observe a change in temperature after the commencement of the task. It is the time between the start of the task and the temperature reaching $3\times$standard deviation above the forehead baseline temperature or below nose baseline temperature. We picked this method, as 99.73% of the data should be within $\pm$ 3 times the standard deviation, hence values outside this range reflects temperature increase/decrease in the forehead and nose respectively.

2. Temperature change ***saturation*** is the time taken to reach saturation in temperature change. This measure describes the time between the onset to the time the temperature lies between $\pm$ $3\times$standard deviation.

3. Temperature change ***offset***, is the time taken after the task is stopped to observe temperature change. We computed based on the time it took between the endof the task and the temperature reaching $3\times$

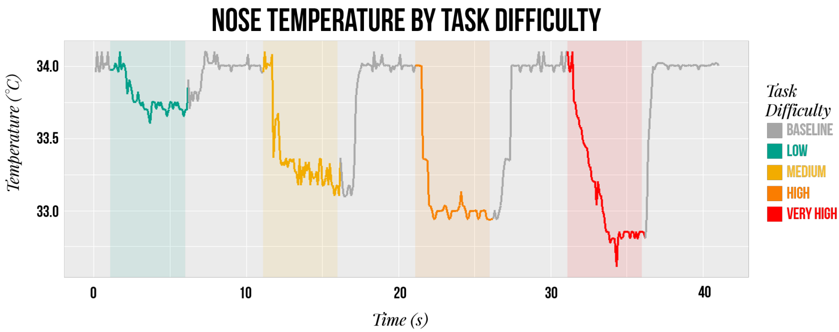



**Figure 5.7:** Nose Temperature Change.

standard deviation below the forehead saturation temperature or above nose saturation temperature.

The baseline temperature for each participant was determined from the relaxing phase. This temperature was compared to the facial temperature during and after the task. We tested the effect of the task duration on the onset, saturation and offset times of the temperature change both in the nose and forehead area. There was no significant difference observed between the task duration and the three metrics.

**Latency in Nose Temperature Change** The onset for the nose temperature decrease was observed after 0.7s (± 0.2s) after the start of the task. It took 3.1s (± 1.2s) to reach saturation temperature . The offset for the nose temperature was observed after 1.1s (±0.5s) the end of the task.

**Latency in Forehead Temperature Change** The onset for the forehead temperature increase was observed after an average of 1.2s (± 0.3s) after the start of the task. It took 2.3s (± 0.9s) to reach saturation of temperature increase. The offset for the forehead temperature was observed after 1.6s (±0.9s) after the task was finished.

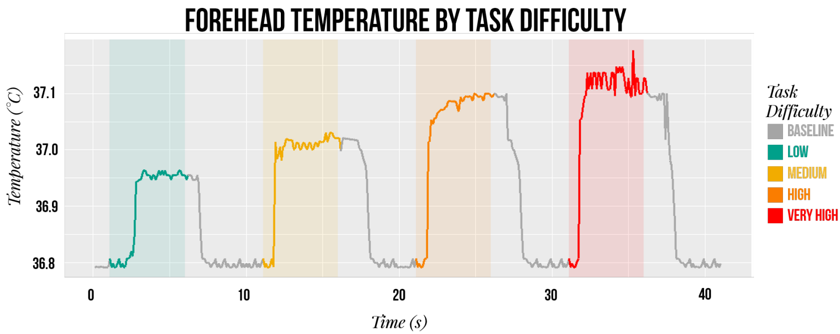



**Figure 5.8:** Forehead Temperature Change.

**Latency in Total Difference Temperature Change** The onset for the differential temperature change was observed after an average of 0.7 ($\pm$ 0.2s) after the start of the task. It took 2.3s ($\pm$ 1.2s) to reach the saturation temperature difference The offset for forehead-nose temperature difference was observed 1.1s ($\pm$0.5s) after the end of the task. As reported in study I, the Stroop test showed a statistically significant difference between the levels and baseline in the three metrics. This was confirmed in the second study.

Thermal variations due to vascular changes were consider as slow compared to other physiological monitoring sensors [129]. Recent research reported a latency of thermal response of 3.8 seconds after the stimuli onset using functional infrared imaging, compared to 3 seconds of GSR [142, 170].

Our findings indicate a response latency of 0.7sec using commercially available, far-infrared thermal imaging. One explanation for the faster response is the camera sensitivity as well as the frame rate, where the camera we used had 0.04K thermal sensitivity and 80 fps, as opposed to the camera used by Kang et al. [129], which had 50 fps and operated in different spectrum. Additionally, we relied on the temperature information of two regions of interest with a simple and real-time ROI extraction approach, which might have influenced the latency of the observed temperature changes.

As presented above, the onset in the forehead was longer than that of the nose 1.2$\pm$0.3 and 0.7$\pm$0.2 seconds respectively. This reflects the fact that there are more vessels affecting the subcutaneous temperature in the nose area than the forehead as reported by Berkovitz et al. [32]. This is also confirmed by the temperature variations, where a temperature change of 0.09° was observed in the

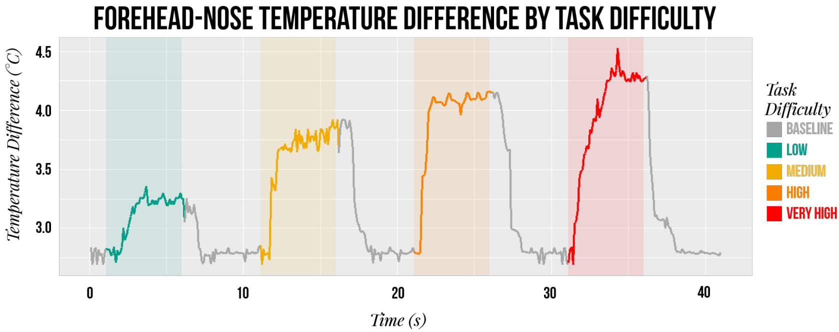



**Figure 5.9:** Forehead-Nose Temperature Change.

forehead as opposed to 0.33° in the nose area, for each level. Our findings from the second study demonstrate the responsiveness of commercial thermal cameras in estimating cognitive load.

### 5.1.5 Discussion and Summary

Our findings validate the correlation of cognitive load and the selected region of interest and the measuring metrics we selected. Fernández-Cuevas et al. and Ioannou et al. [117, 80] summarized and presented how facial temperature and region of interest vary with different mental states. However, there were no states (e.g. stress, guilt, joy, etc.) that correlated with an decrease in the nose and an increase in the forehead temperature. For instance, fear was correlated with decrease in both nose and forehead temperatures [117] and stress was correlated with variation in the nose temperature [117, 129]. Other works identified stress as an equation of the difference of the temperature of the nose and forehead [131] with a specific reading values between 34 and 36 degrees, rather than the total temperature change.

Our results show that the order of magnitude of the temperature changes are large enough to be detected by commercial sensors. For example, the FLIR One, a smartphone-compatible thermal camera and one of the most affordable devices currently in the market, is capable of detecting temperature changes of 0.1 degrees Celsius and can hence be used to detect cognitive load.

In this work we described our approach to unobtrusively derive users' cognitive load based on thermal imaging. Therefore, we investigated the effects of four different task intensity levels on facial temperature changes. We implemented a system capable of monitoring forehead and nose temperature to estimate current cognitive load levels through a novel metric based on the difference between forehead and nose temperature.

Thermal imaging operating in FIR provides novel avenues for studying users' cognitive states. We observed substantial changes in facial temperatures upon the activation of the ANS due to a stimulus. While the nose temperature—reflecting the vasoconstriction limiting the blood flow to the surface i.e. skin—decreases with rising workloads, in parallel, the temperature on the forehead increases as muscular activity leads to metabolic increases and increased blood flow in the underlying vessels. Based on these observations, we proposed a novel unobtrusive technique for estimating and quantifying cognitive load and possibly other affective states. In addition, we investigated the latency of temperature change and the ability of thermal cameras to capture those changes. We found an average latency of 0.7 $\pm$0.2 seconds.

Therefore, our system was able to unobtrusively estimate changes in cognitive load in close to real-time. The exploration of content types gives rise to thermal-based activity tracking, which can empower new applications in the field of cognition-aware computing. Thermal imaging techniques, for example, can be applied in classroom settings with multiple students being monitored in real-time to estimate cognitive load levels and assess current difficulty of content. It could also be used in assistive systems in a production environment, where the worker is monitored unobtrusively without interrupting their work flow to estimate the current difficulty of the task in hand. Additionally, our proposed system could be utilized in usability testing to help identify user interface features that increase cognitive load. Awareness of cognitive demand allows systems to dynamically adapt to users' current cognitive capacities and either reduce task difficulty to prevent frustration or add complexity to sustain interest and productivity.

## 5.2 AttenTCam: Classifying Attention Type

*This section is based on the following publication:*

- Y. Abdelrahman, A. Khan, J. Newn, E. Velloso, S. Safwat, J. Bailey, A. Bulling, F. Vetere, and A. Schmidt. Classifying attention types with thermal imaging and eye tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technology*, Sept. 2019

Building on our findings from the previous section, we further utilize thermal imaging augmented with other unobtrusive sensors, in particular an eye tracker, to classify attention types. Our approach exploits the fact that each attention type requires different cognitive resources [164]. This, in turn, influences how the blood flows through the face, which is then reflected in variations in the skin temperature [191, 12]) as a reflection of the activation of the ANS [73]. In short, we are investigating the leverage of the correlation between cognitive state and facial temperature combined with eye movements to differentiate between attention types.

A common, though often incorrect, hidden assumption underlying how we currently design interactive systems is that, during the interaction, the user focuses all of their attention on the interaction with the system. As a consequence, considerable effort in the research and development of ubiquitous computing systems has been placed on supporting users while they perform single-focus tasks (e.g., [84]). However, given the multitude of devices and applications constantly fighting for users' attention through interruptions and notifications, single-focus interactions are the exception rather than the rule [249, 263].

This phenomenon has led economists to frame the problem in terms of an *attention economy*, where attentional resources are the currency and actors are competing for consumers' attention [76, 55]. A fundamental concept in this idea is that, similar to other economic resources, *attention is a limited resource*. Further, attention is strongly influenced both by *internal stimuli* (e.g. remembering where you left your keys causes your attention to shift, or feeling motivated to read a book leads to a more focused reading experience) and *external stimuli* (e.g. hearing a dog bark behind you causes you to turn around or writing an essay while being observed by your teacher keeps your eyes on the exercise at hand). Therefore, the context around the interface affects how much attention is paid to the interaction.

To further complicate matters, attention itself is a complex concept, one that even psychologists struggle to conceptualize [238]. Early studies suggested that there are several levels of attention instead of a unitary one, due to its complex nature involving memory, behavior and consciousness [132, 171, 237, 164, 200]. One model that emerged from this literature is Sohlberg and Mateer's *Clinical Model of Attention* [230]. This hierarchical model discriminates between people's ability to maintain attention towards a single stimulus (*sustained* and *focused* attention); to switch attention between different stimuli (*alternating* attention); to pay attention to one stimulus while inhibiting others (*selective* attention), and to pay attention to multiple stimuli simultaneously (*divided* attention). This model highlights two challenges: quantifying attention (how much attention) and qualifying the nature of attention (what type of attention).

Prior work on attention has shown that our well-being is tied strongly to our ability to manage our attention successfully, for example, we know that multitasking hinders performance [152]. Such known issues create opportunities to design interactive systems that monitor and actively help users to manage their attention. The vision of *pervasive attentive user interfaces* encapsulates this well, stating that interfaces could "adapt the amount and type of information based on [users'] current attentional capacity, thereby simultaneously optimizing for information throughput and subtlety. Future interfaces could trade-off information importance with users' current interruptibility level and time the delivery of information appropriately" [40].

To realize this vision, interfaces that attempt to leverage the users' attention must accomplish two tasks: (1) *identify* the locus of attention and (2) *characterize* the nature of the current attentional state. While the locus of attention is typically considered to be equivalent to gaze direction, this is not always the case due to the diverse nature of attention orienting, which is classified as overt or covert [285]. In overt attention, the person selectively attends to a source of information by moving their eyes to point in that direction [199]. However, humans do not necessarily direct their eyes towards their area of focus. During covert attention, a corresponding shift in attention is not followed by a corresponding shift in gaze direction [74], e.g., when a person has a conversation with a friend while looking at their mobile phone [164], or when eavesdropping on a conversation while typing up an email in an open office environment. Therefore, even though eye-tracking data can be very informative, it is essential to understand the limits of the gaze point as a sole representation of the locus of attention.

In this work, we address the limits of eye-tracking for attention detection by proposing its combination with thermal imaging in order to classify the various attention types by stated in Sohlberg and Matter's *Clinical Model of Attention*. The Clinical Model describes attention as a model based on the degree of focus, consisting of lower fundamental levels and higher levels [230]. The lower level includes *focused* and *sustained* attention, while the higher levels includes *selective*, *alternating*, and *divided* attention [230, 132, 164]. The Clinical Model further describes attention as a multidimensional cognitive capacity, which means that attentive tasks need different levels of cognitive load to be achieved [164]. The findings of recent work in HCI demonstrated the ability to unobtrusively quantify cognitive load using thermal imaging and temperature sensors [13, 240, 289]. Our work, therefore, builds on the ability to use thermal input as a method of measuring different levels of cognitive loads and the knowledge that these different attention types require different cognitive capacities. By combining this concept with the ability to detect overt attention reliably well through eye-tracking, we explore the novel combination of eye-tracking and thermal imaging for attention classification. To explore this combination, we collected a dataset in a user study designed to elicit different attention types using different stimulus modalities, in controlled and (semi-)naturalistic tasks. We build on the opportunity that eye-tracking can reveal the locus of attention, and thermal imaging can give us an estimate of cognitive load. Together, this allows us to paint a better picture of users' attentional state. We hypothesize that by combining these modalities, we are able to classify different attention types according to the Clinical Model.

### 5.2.1 Related Work

In the past decades, many scientific fields have been interested in understanding the processes behind human attention, from its measurement to its modeling. A pre-condition for this is the ability to sense and characterize attentional states in near real-time, and prior work has explored the use of various sensors and algorithms in attempts to achieve this. In this section, we discuss the background theory of attention, technology-based approaches for sensing attention, and existing algorithms for classifying attention.

The vast body of research on theories of attention can be split loosely into theories of focused attention and theories of divided attention, with few studies attempting to bridge the gap between the two (e.g., [182]). Whereas theories of focused attention are grounded on visual selection and unintentional processing, theories of divided attention revolve around the issue of capacity limits [77, 126].

These differences in theoretical grounding have led to the evolution of different attention models in the field of psychology. In our work, we employ Sohlberg and Mateer's *Clinical Model of Attention* [230] as it has been deemed to be one of the most comprehensive models [25]. The Clinical Model describes attention as a model based on the degree of focus, consisting of lower fundamental levels and higher levels [230]. The lower level includes *focused* and *sustained* attention, while the higher levels includes *selective*, *alternating*, and *divided* attention [230, 132, 164]. In other words, attentive tasks need different levels of cognitive load to be achieved [164]. The attention types introduced in the model are:

**Focused attention.** The ability to respond discretely to specific visual, auditory, or tactile stimuli.

**Sustained Attention.** The brain can discretely respond to specific auditory, tactile, or visual stimuli for a prolonged period. Reading a book in a deeply focused state is an example of sustained attention.

**Alternating Attention.** Happens when we switch focus from one task to another or from one part of the task to another, regardless of different cognitive demands between them. Examples include: listening to a lecture while taking notes, or reading a recipe while cooking.

**Selective Attention.** The ability of the brain to focus on a specific stimulus while inhibiting others. A prime example of selective attention is called the *Cocktail Party Effect* [144], which describes our ability to selectively attend to the voice of one person while minimizing other voices and noise.

**Divided Attention.** The brain divides its attention between different stimuli simultaneously. Examples include: playing a mobile game while having a conversation or, one that we do not recommend, texting while driving.

## Current Approaches to Classify Attention

A crucial step in building attentive systems lies in the ability to quantify users' attentional states. However, as changes in these states happen inside users' minds, we can only measure attention indirectly through users' behaviors and physiological signals, leading to the development of technologies potentially offering insights about the users' attention states. These technologies vary in their levels of obtrusiveness.

Previous works have explored a variety of sensors for measuring attentional states, including electroencephalography (EEG) [2, 158, 154, 140], electrooculography (EOG) [140], electrocardiography (ECG) [46], and electromyography (EMG) [204]. These sensors have been shown to provide high accuracy in recognizing states but are obtrusive in nature (requiring users to wear a device or have electrodes attached to their skin), and therefore cumbersome for daily use. For instance, Liu et al. [158] were able to distinguish between attentive and inattentive states with an accuracy of 76.82% using EEG but required the placement of electrodes on participants' heads. On the other hand, researchers have employed less unobtrusive approaches such as functional Magnetic Resonance Imaging (fMRI), commonly used to reveal aberrant brain activity, to measure attentional states [160, 108, 173]. For example, Moisala et al. [173] measured human brain activity during single-tasking and dual-tasking using fMRI, looking for activation in the medial and lateral frontal regions of the brain. Their results highlight the relationship between different attentional demands and levels of brain activity associated with sustained and divided attention. Though able to show differences in attention states, fMRI remains impractical for daily use, in terms of costs and practicality.

Recent work has explored unobtrusive contactless sensing approaches, including eye tracking and temperature sensors. Eye tracking is a common technique to investigate visual attention as we tend to fixate on objects that have drawn our attention or relevant to the task that we are attending to [164, 179, 178]. Through our visual attention, we only 'see' what we are paying attention to, as our cognitive system allocates sufficient resources for visual processing to avoid overloading. Because we receive a large amount of information through our eyes, this mechanism helps us to manage what gets processed. Eye movements are an important part of visual attention and are primarily comprised of fixations (stationary phase) and saccades (rapid, ballistic eye movements phase). Previous works have long explored how eye movement features can help uncover psychological states and recognize activities [246, 233]. Eye tracking is a powerful tool for understanding human attention as it can measure both the frequency of eye movements and the location of the gaze point [164]. While researchers often use gaze point as a proxy for the locus of attention, this is not always the case due to the diverse nature of attention orienting—classified as overt or covert [285]. Therefore, even though eye tracking data can be very informative, it is essential to understand the limits of the gaze point as a sole representation of the locus of attention.

Thermal imaging and temperature sensors have been explored as a means of understanding users' mental states [13, 240, 116, 151, 211], for instance, thermal imaging has been used to detect several states including stress, guilt, fear [117]. Our work, however, builds specifically on *Cognitive Heat* [13] and Zhou et al.'s *Cognitive Aid* [289], which demonstrate the relationship between facial temperature and cognitive load estimation, in which the authors employ the use of thermal imaging as a way to unobtrusively detect changes in cognitive load in real-time. To elaborate, the authors found substantial changes in facial temperatures upon the activation of the ANS when exposed to the stimulus, specifically between the nose and forehead regions. This seminal work gave rise to developing thermal-based activity tracking, which further facilitates new applications in the field of cognition-aware computing. Wearable variations have also been developed using the same concept. For example, Tag et al. [240] presented early work on the use of facial temperature to measure attention; demonstrating the ability to measure attention using IR temperature sensors. However, their focus was attention level rather than type. Similarly, Zhou2019cognitiveaid explored the use of thermal sensors to detect mental workload, demonstrating the ability of such sensors to detect when a user is currently performing a task.

The variety of sensors discussed above opened the opportunity to use machine learning techniques to classify users' mental states and to build systems that adapt to these states [221]. Whereas there have been initial efforts to use machine learning to classify attention primarily into attentive and non-attentive states with a maximum accuracy of 93.10%, no work has attempted to classify attention according to the four types of attention outlined in Sohlberg and Mateer's Clinical Model of Attention [230].

## Summary & Research Direction

In summary, there are two clear limitations from the existing literature on recognizing attention types. First, the sensors employed for measuring attention tend to be obtrusive and therefore not appropriate for the development of interactive systems. Second, works to date employed models that oversimplify attentional processes, as a binary variable or as a one-dimension continuous signal.

In this work, we address this research gap by using the combination of two unobtrusive sensors—*thermal imaging* and *remote eye tracking*, from which we can build classifiers for recognizing the four types of attention outlined in Sohlberg and Mateer's *Clinical Model of Attention* [230]. To our knowledge, this is the first work that has attempted to differentiate between four attention types (*sustained*, *alternating*, *selective*, *divided*). Our combined approach exploits the fact that each attention type requires different cognitive resources [164] and visual direction [82].

To elaborate, our novel approach leverage two ways in which attentive states are manifested in our physiology to measure and classify attention effectively. First, attention is related to the allocation of cognitive resources [164]. This process strongly correlates with changes in the blood flow, which is reflected in changes in the temperature distribution in our skin [229, 223, 13]. In a earlier work, we explored the use of thermal imaging to measure cognitive load, in which we relied on how the activation in the ANS due to an increase in cognitive load is reflected in the facial temperature. Using the same ideology, we hypothesize that changes in attentive states will also lead to a change in cognitive load levels that are observable in facial temperature patterns measured with a thermal camera. Our hypothesis is built upon the fact that different attentive states require different levels of cognitive load [164]. Informed by the literature, we estimate cognitive load using the nose-forehead differential [117, 13]. Also, we explored the effect of different attentive states on the user's cheeks, as previous work [117] highlighted the usage of cheeks as state indicator. Second, when engaging in overt attention, the gaze point—which we can easily measure with eye-tracking—is a reasonable estimate of the locus of attention [82]. Further, low-level statistical features of eye movements are also indicative of cognitive load levels, which can be useful for an attention classifier (e.g. [287]).

These physiological properties present an opportunity for the design of pervasive attentive user interfaces. Both eye movements and facial temperature patterns can be unobtrusively captured with remote eye trackers and thermal imaging cameras, particularly considering that the face is the most often exposed part of the user's body. Moreover, recent advances in both eye-tracking and thermal imaging have made it cheaper and more accessible than ever to capture this information without the need to augment the user, but rather the environment.

In the following sections, we present the data collection with a detailed description of the tasks used for attention elicitation (Section 5.2.2). We hypothesize that the higher-level attention types (selective and divided) will result in a more significant temperature difference. To explore this hypothesis, we conducted a user study to elicit attention types using the combination of audio and video stimuli, while recording the gaze and thermal data (Section 5.2.2). We then present our methodological approach to analyze the collected data set, including statistical analysis, feature extraction, and classification. In Section 5.2.3, we report the results from different classification approaches (user-dependent, user-independent (condition dependent) and user-independent (condition independent), showing the applicability of thermal cameras and eye tracker as unobtrusive sensors to classify attention. Lastly, we discuss how the findings of our work can be applied and present directions in the future work section.

### 5.2.2 Attention Type Classification

Our goal is to build a classifier that is able to distinguish between attention types based on facial thermal imaging and eye-tracking data. To train and evaluate this classifier, we collected a dataset in which we recorded the eye movements and the temperature of facial features (nose, forehead, left cheek, and right cheek) of 22 participants as they completed tasks designed to elicit four types of attentional states. The tasks were inspired by the literature on *attention* in psychology. We used a repeated-measures design, where all participants performed four sets of tasks with different stimuli. We counterbalanced the order of the tasks. We created variations of each stimulus to elicit four types of attention in the *Clinical Model of Attention* — *sustained*, *selective*, *alternating* and *divided*, for a total of 16 tasks (4 attention types $\times$ 4 types of stimuli). We did not include focused attention, as we are interested in the attention over prolonged periods. Further, we included a baseline task at the beginning of the experiment.

#### *Tasks*

We used a combination of tasks from the attention elicitation literature and developed a series of tasks to elicit different attention types starting with Stroop conditions as a reference task, followed by more naturalistic tasks that involved a combination of visual- and audio-based stimuli. For the baseline task, we asked the participants to relax while listening to white noise. We used the baseline task to capture and record the participants' temperatures at rest, which serves as a point of comparison with the other tasks [13]. Figure 5.10 illustrates the remaining tasks used in the study. We published a playlist with the stimuli online[33], for reproducibility purposes. We displayed the tasks in full screen for 3 minutes, and conditions without a visual stimulus contained a white background. For consistency, we primarily used selected TED Talks[34] for the content of the tasks. In audio-based tasks, we extracted the audio from the videos, while in the visual tasks, we used the transcripts of the talks.

[33] `https://bit.ly/2LyZWay`

[34] https://www.ted.com/talks

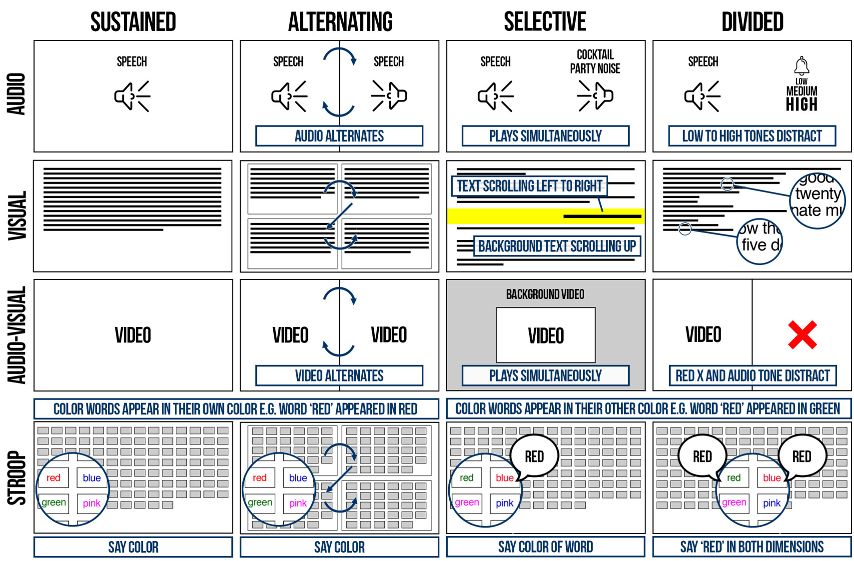



**Figure 5.10:** Conditions to stimulate the four different attention types.

### Stroop Tasks

The Stroop test is a classic Psychology task for eliciting selective attention [236, 147]. In the typical test, users are asked to name the color of the font in which words are written. The difficulty of the task lies in the fact that the words displayed correspond to a different color to the one in which they are colored while the user *selectively attends* to the color of the font. For example, in the classic experimental task, the word 'RED' would be colored in blue, and the participant must reply 'Blue' while ignoring the fact that the word itself corresponds to a different color. For our study, we created three variations of the Stroop test to elicit the remaining attention types, described below:

*Sustained Stroop:* We first created a simplified variation of the Stroop test to elicit sustained attention, where we retained a single source of information. We showed color names written in their own color and asked participants to read it aloud. For example, the word 'green' would appear colored in green, and participants were asked to say 'green'. This effectively removed the challenge of the task allowing the participant to focus on reading the words, therefore maintaining sustained attention.

*Alternating Stroop:* In this variation of the Stroop test, the display was split into two halves. Each half had a sustained Stroop test variation, and participants had to alternate between the two halves, spending 45 seconds in each half.

*Selective Stroop:* We used the original Stroop test [236] as the selective Stroop test where text and color are presented differently. For example, the word 'green' would appear colored in blue, and the participant had to say 'blue.' Participants, therefore, have to 'selectively' choose between the two.

*Divided Stroop:* We reused the Stroop test variation introduced by Eidels et al. [68] to elicit divided attention where participants are directed to attend to both word and color. The task included all four combinations of the words, RED and GREEN, and the ink colors, red and green. The participants are asked to respond to 'redness' in the Stroop stimulus, regardless of whether the 'redness' comes in the word (RED), the color (red), or both (RED in red). Hence, the participant must attend to the color and to the word (i.e. divide attention across the Stroop stimulus components).

**Audio Tasks**

*Sustained Audio:* This task used a single audio file of a TED talk speech to which participants were asked to listen attentively.

*Alternating Audio:* To simulate a group conversation, we used two audio sources, which alternated between being on and off every 45 sec. We used the same topic to mimic the real-life example of a group conversation.

*Selective Audio:* We simulated the *Cocktail Party effect* [144], where we combined audio of a speaker with the audio of a cocktail party. Participants were asked to attend to the speaker selectively.

*Divided Audio:* Inspired by Gardiner et al. [87] to elicit divided attention, participants were asked to listen to a presentation talk while listening to and reporting high, low, or medium tone sequences by saying the tone level out loud.

**Visual Tasks**

*Sustained Visual:* A single panel of text was displayed, and participants were asked to read the text as it appeared.

*Alternating Visual:* We divided the screen into two panels, which we further subdivided into two parts. The text first appeared in the left panel for 45 sec and then alternated to the right panel, the text then alternated back to the bottom half of the first panel and lastly alternated to the bottom half of right panel. Participants are asked to read the text displayed in the active panel.

*Selective Visual:* A stream of text was displayed in a highlighted region in the middle section of the screen from left to right. A stream of text also flowed upwards in the background. Participants were asked to read the text in the highlighted region selectively. This task was inspired by a news ticker (also called slide) that typically appears at the bottom of TV channels.

*Divided Visual:* In this task, we augmented numbers into a text transcript. We asked participants to read the text while performing mental addition on the numeric values that appeared in the text, e.g. twenty, five, etc. This forced the participants to divide their attention between the text itself and the mental arithmetic task.

**Audio-Visual Tasks**

*Sustained Audio-Visual:* For this task, we had a single video running from a selected top TED talk.

*Alternating Audio-Visual:* Similar to the Alternating Visual task, the screen was divided into two panels and two videos played alternatively in the two panels. The first video played for 45 seconds and alternated to the second panel. This alternating process repeats twice.

*Selective Audio-Visual:* Two videos were displayed, one embedded in the other, as shown in Figure 5.10. Participants were asked to selectively attend to the video with the talk that was displayed in the middle of the screen. The larger video acted as a cocktail party like noise [87].

*Divided Audio-Visual:* Inspired by Gardiner et al. [87] to elicit divided attention and similar to the Divided Audio task, participants were asked to watch a video while listening to and reporting high, low, or medium tone sequences. Additionally, we added an appearing “X” in Red, and the user was asked to say "X" out loud when the symbol appears to elicit attention on divided audio-visual type stimuli.


**Figure 5.11:** Study setup.

### *Experimental Setup*

Figure 5.11-Left illustrates our experimental setup, consisting of a commercial Tobii EyeX eye tracker[35] operating with frequency of approximately 55 Hz, connected via USB. The eye tracker provided the gaze *x*- and *y*-coordinates on the screen. We attached the eye tracker to a 24" screen and placed an Optris PI450 thermal camera[36] mounted on a tripod 1m away from the participant behind the screen. The camera has an optical resolution of 382×288 pixels, has a frame rate of 80 Hz, and measures temperatures between -20°C and 900°C, with a thermal sensitivity of 0.04°C. Further, the camera captured wavelengths in the spectral range between 7.5$\mu$m and 13$\mu$m with a 38°× 29° field of view. The output of the camera encodes temperature information with 16-bit color values. Further, we developed a system to display the stimuli (tasks) for each test in a counterbalanced order using Latin square that records both streams of data. The Optris PI connect software [37] used with the camera has a built-in annotation function, using the so-called measure areas of 10×20 pixels. We annotated the regions of interest including forehead, nose and cheeks, as depicted in Figure 5.11-Right. Additionally, the Optris PI connect has a built-in save option, that stores the mean temperature values of the annotated regions in CSV files.

### *Participants & Procedure*

We recruited a total of 24 participants, and discarded 2 participants due problem with eye-tracking calibration. The remaining 22 participants in our final data

---

[35] `https://tobiigaming.com` (recent firmware upgrade enabled increased frequency to 70Hz)

[36] `http://www.optris.com/thermal-imager-pi450`

[37] `https://www.optris.com/`

set consisted of 14 Males and 8 Females with an average age of 20.45 years ($SD = 1.14$), recruited through university mailing lists. Upon arrival, participants were asked to sign a consent form and were informed about the aim of the study. We first asked participants to relax for 5 minutes while listening to white noise (relaxing sound of ocean waves) as the baseline task. This allowed us to collect their physiological data in a state of relaxation. Following, we presented the different tasks, 16 tasks in total for 3 minutes each. We explained each task to the participants before starting the task. The order of the tasks was counterbalanced using Latin squares. After each task, we asked participants to complete a NASA-TLX [101] questionnaire to assess the perceived cognitive load. The study lasted approximately 85 minutes ($SD = 10.25$). During the entire experiment, we recorded the facial temperature and eye gaze coordinates of the participant. The study was recorded using an RGB video camera (further described in the next section). We maintained the room temperature at 23°C, and participants were compensated with 10 EUR upon completion.

## Method

In this section, we describe our step-by-step process in which we use to evaluate the combination of thermal imaging and eye-tracking for attention classification. First, we statistically analyzed the results to evaluate objective and subjective measures. Second, we extracted the features required for classification. Third, we built and tested different classifiers based on these features. We then measured the best performing classification model on our different classifiers before diving down into the performance of the combination.

### *Statistical Analysis*

To validate our attention elicitation, we analyzed the effect of the attention types on both the subjective cognitive load from the NASA-TLX reported by the participants and the cognitive load inferred from the recorded facial temperature. We used three metrics as our dependent variables: the NASA-TLX score, forehead-nasal temperature, and cheeks temperature (detailed in Section 5.2.3). We statistically analyzed the data using a repeated measures ANOVA (with Greenhouse-Geisser correction if sphericity was violated). This was followed by posthoc pairwise comparisons using Bonferroni-corrected t-tests.

Feature Extraction To train our classifiers, we derived a feature set (14 features) that best describe the various attention types from both the gaze and thermal data

| Type | Subcategory | Feature |
|---|---|---|
| Gaze | Stimulus-dependent | Number of gaze transitions between pairs of Area of Interest (AOI). AOI where maximum fixation lies in a window of 45 sec. |
| Gaze | Stimulus-independent | Number of fixations.Mean fixation duration. |
| Thermal | Both | Mean forehead and nose temperature difference from the baseline. Mean temperature change in the cheeks from the baseline. |

**Table 5.4:** Selected feature set used for classification.

(see Table 5.2.2 below). Below, we explain our reasoning behind our choices of features. The details of how we trained our classifier can be found in Section 5.2.2.

Gaze Features Stimulus-dependent features are those that involve the knowledge of the AOI of the interface, whereas stimulus-independent features are statistical measures computed from eye movements. We pre-processed the gaze data by removing outliers and by clustering gaze points into fixations. We identified fixations using the Dispersion-Threshold Identification algorithm [214], as it produces accurate results in real-time using only two parameters, dispersion, and duration threshold (set to 20 and 100, respectively). From this data, we computed low-level statistical features, such as the number of fixations and mean fixation duration, as shown in Table 5.2.2.

As a representative example, Figure 5.15 shows the gaze plots for all combinations of task and attention type for one participant. For our purposes, the meaning of the area under the gaze point in regards to the task at hand is an important factor in determining the attention state. For example, consider a system that monitors a student while they watch a video lecture. Two similar fixation patterns will be indicative of attentive or inattentive states depending on whether it falls inside or outside the video player. Therefore, as suggested by Toker:2013, in addition to the stimulus-independent features, it is important also to compute stimulus-dependent features that encode the meaning behind different AOI. Hence, we divided the task interface into different numbers of AOI depending on the stimuli (see Figure 5.12). The stimuli-dependent features extracted were the number of fixations in an AOI for every 45 seconds and the number of gaze transitions between pairs of AOI. To compute the gaze pattern, we used the number of fixations in each AOI to identify the area with the highest number of fixations fixation every 45 seconds. Though in our experimental setup, we manually created the AOI, in a real system implementation, they could be set by the UI implementation framework used for its development.

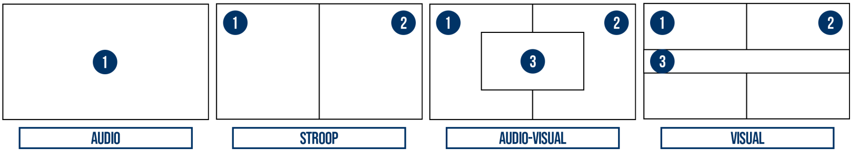



**Figure 5.12:** Area of Interest (AOI).

### *Thermal Features*

Previous works on thermal imaging for users' mental state detection [13, 116, 240], build upon the fact that changes in the internal states influence the blood flow [229]. Because our blood carries heat, changes in the blood flow influence our skin temperature [116, 117, 223, 229]. Therefore, monitoring changes in facial temperature can give us an insight into the changes in mental states. Researchers explored multiple regions of interest in the human body e.g., mouth, nose, and hand [117]. In particular, the face showed potential in detecting changes in states, as it is exposed and easy to capture by thermal cameras. Furthermore, it has a thin tissue layer, making temperature changes more observable. Therefore, in this work we explore how facial temperature fluctuations can give us an insight into changes in cognitive load caused by the experienced different attention types. We computed the temperature difference of the cheeks, forehead, and nose from their mean baseline temperature, similar to previous works [13, 116, 240].

## Classification Approach

The goal of our classifier is to map a feature vector computed from a window of data to one of four classes corresponding to the type of attention the user was engaged as per the Clinical Model of Attention. To do this, we first built a *user-dependent, condition- independent* classifier, which was trained on the data from the same participant but different condition on which it was evaluated. This was followed by a *user-independent* classifier which was trained on the data from different participants on which it was evaluated. We then further evaluated the user-independent classifier in two ways—*condition-dependent* and *condition-independent*. To put simply, we trained the *condition dependent* variant on the data from other participants in the same condition (leave-one-out-cross-validation on participant), while the *condition-independent* classifier is trained on a different set of conditions and users to the dataset on which it was evaluated, e.g., trained on the Stroop, Audio and Visual datasets and evaluated on the Audio-Visual (leave-one-out-cross validation on participant and task). We provide more details on the three distinct classifiers in the remainder of this section. As different classification models will generate different levels of performance, we compared three different classifiers for all three classifiers: Support Vector Machines (SVM), K-Nearest Neighbour (KNN) and Logistic Regression (LR). For the SVM classification model, we used the two hyper-parameters C=5 and gamma=0.01 with RBF kernel, while the KNN model was trained with k=5 neighbours. We used the scikit-learn package[38] for machine learning in Python for feature extraction and classification and PyCharm[39] as a development environment.

### *User-Dependent Classifier*

We built a user-dependent, condition-independent classifier by training the data on the same participant but different condition for the four tasks. This allows us to evaluate the performance of our approach of a system that is trained on its own user (e.g. by having a calibration phase). To do this, we trained and evaluated the classifier 22 times, using all 14 features, each time for a specific participant for the remaining conditions. For example, we trained the classifier on the data of a participant of Stroop, Visual and Audio task and evaluated the classifier on the data of the same participant but the Audio-Visual task).

### *User-Independent Classifiers*

User-dependent classifier can potentially be optimistic, we next built a user-independent classifier. Being independent of the user, we can obtain a more robust and generalized classifier. We further split the user independent classifier in *condition-dependent* and *condition-independent* variants.

### *Condition-Dependent*

We evaluated the classification performance of the condition-dependent classifier on the data from the same condition on which it was trained, but from a different participant. We conducted separate evaluations for each task (Stroop, Audio, Audio-Visual, Visual), building and evaluating the classifiers using leave-one-participant-out cross-validation. We trained the classifier 22 times, each time training on the data of 21 participants and evaluating it on the remaining one participant.

[38] `https://scikit-learn.org/stable/`

[39] `https://www.jetbrains.com/pycharm/`

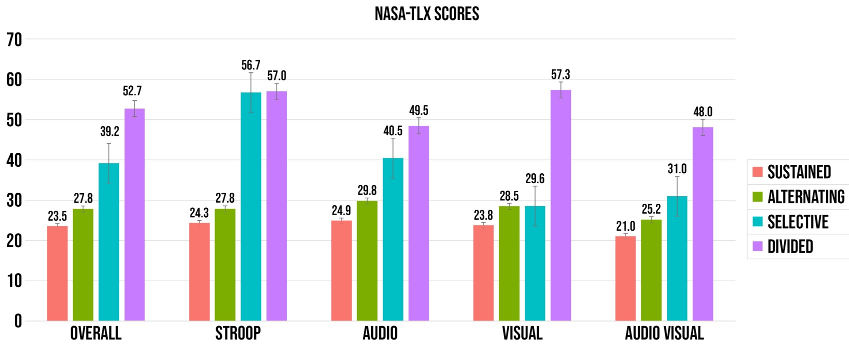



**Figure 5.13:** The average cognitive load perceived by the participants.

*Condition-Independent*

We evaluated the condition independent classifier by training it 22 times using leave-one-participant-out cross-validation four times, one for each condition. Each time, we trained it on the data of the 21 participants for three conditions and evaluated it on the data of the fourth condition from the last participant. The reported results, in the next sections, are averaged by participant but split by the task on which it was evaluated.

## 5.2.3 Results

*Statistical Analysis*

Below we present the effect of the attention types of the different conditions on the facial temperature as opposed to the baseline.

*Cognitive Load: NASA-TLX*

To confirm that each attention type requires different cognitive resources [164], we first analyzed the effect of the different attention types on the reported cognitive load via the NASA-TLX. We tested the effect of the different attention types from different conditions on the overall cognitive load.

Condition-Independent NASA-TLX: We first analyzed the mean NASA-TLX SCORE from all conditions (Stroop, Audio, Visual, and Audio-visual) for the four ATTENTION TYPES. As depicted in Figure 5.13, the sustained attention had the lowest load with an average score of 23.50 (SD = 12.22), followed by the alternating attention with an average of 27.81 (SD = 14.59), selective attention with an average of 39.17 (SD = 15.88) and the highest was divided attention with an average score of 52.72 (SD = 18.49). We tested the effect of the ATTENTION TYPE (4 types) on the overall NASA-TLX SCORE with a one-way ANOVA. Mauchly's test showed a violation of sphericity against difficulty ($0.29, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.65$) values. We found a significant large effect of ATTENTION TYPE on the NASA-TLX SCORE ($F_{2.09,41.72} = 72.29, p < .001, ges = 0.38$). Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$).

Stroop NASA-TLX: We further analyzed the NASA-TLX score from the Stroop condition. The sustained attention had the lowest load with an average score of 24.32 (SD = 18.23), followed by the alternating attention with an average of 27.84 (SD = 22.51), selective attention with an average of 56.70 (SD = 24.03) and the highest was divided attention with an average score of 57.01 (SD = 22.16). We tested the effect of the attention Type (4 types) on the NASA-TLX Score with a one-way ANOVA. Mauchly's test showed a violation of sphericity against difficulty ($0.45, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.73$) values. We found a significant large effect of attention type on the NASA-TLX score ($F_{2.30,46.02} = 30.16, p < .001, ges = 0.34$). Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, and between the selective and the divided attention.

Audio NASA-TLX: In the audio condition, the sustained attention had the lowest load with an average score of 24.89 (SD = 18.23), followed by the alternating attention with an average of 29.78 (SD = 17.03), selective attention with an average of 40.45 (SD = 16.63) and the highest was divided attention with an average score of 48.45 (SD = 20.57). ANOVA revealed a significant effect of attention type on the NASA-TLX score ($F_{3,60} = 13.04, p < .001, ges = .21$). Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, and between the selective and the divided attention.

Visual NASA-TLX: Again, the sustained attention had the lowest load with an average score of 23.77 (SD = 13.24), followed by the alternating attention with an average of 28.48 (SD = 17.56), selective attention with an average of 28.55 (SD = 15.72) and the highest was divided attention with an average score of 57.34 (SD = 18.18). Mauchly's test showed a violation of sphericity against difficulty

$(0.51, p < .05)$, so we report Greenhouse-Geisser-corrected ($GGe = 0.75$) values. ANOVA revealed a significant effect of attention type on the NASA-TLX score ($F_{3,60} = 57.96, p < .001, ges = .41$). However, Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, between the sustained and the selective, and between the alternating and the selective attention.

Audio Visual NASA-TLX: Lastly, for the audio-visual condition, the sustained attention had the lowest load with an average score of 21.02 (SD = 13.29), followed by the alternating attention with an average of 25.15 (SD = 14.78), selective attention with an average of 30.98 (SD = 19.88) and the highest was divided attention with an average score of 48.07 (SD = 19.80). ANOVA revealed a significant effect of ATTENTIO TYPE on the NASA-TLX SCORE ($F_{3,60} = 24.88, p < .001, ges = .27$). Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, and between the alternating and the selective attention.

In summary, the different attention types exhibited different NASA-TLX, where sustained attention showed the lowest NASA-TLX score, followed by the alternating, then selective, and the highest score was observed in the divided attention. Additionally, we found a significant difference in the NASA-TLX.

### *Cognitive Load: Facial Temperature*

Informed by the literature, cognitive load could be assessed by monitoring the facial temperature [13, 289], namely the difference between forehead and nose temperature (difference to the baseline). Other work [117], also investigated the temperature of the cheeks. In this work, we analyzed the effect of the attention type on the Forehead-Nasal temperature and the Cheeks temperatures.

### *Effect of Stroop tasks on Facial Temperature*

**Forehead-Nose** We tested the effect on the total change in the forehead and nose temperature. Mauchly's test showed a violation of sphericity against attention type $(0.47, p < .05)$, so we report Greenhouse-Geisser-corrected ($GGe = 0.55$) values. A large significant effect of attention type on the Forehead-Nose difference ($F_{2.3,46.4} = 39.22, p < .001, ges = 0.63$) was found. Bonferroni-corrected post-hoc tests shows significant differences between all

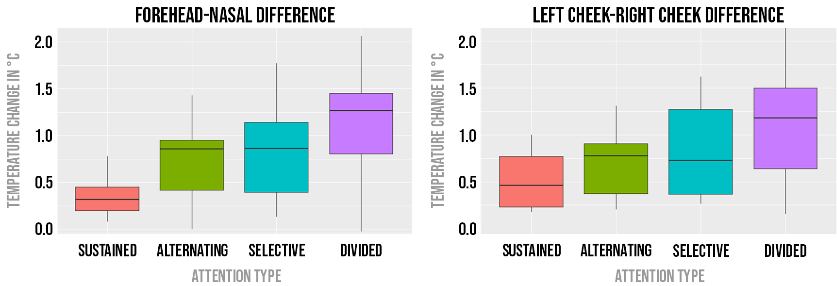



**Figure 5.14:** Mean temperature change.

types of attention ($p < .05$), except between the sustained and the alternating, and between the selective and divided attention.

**Cheeks** For the cheeks temperature, Mauchly's test showed a violation of sphericity against attention type ($0.10, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.54$) values. A large significant effect of attention type on the cheeks temperature increase ($F_{1.6,33.9} = 50.36, p < .001, ges = 0.49$) was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention, except between the sustained and alternating attention.

### *Effect of Audio task on Facial Temperature*

**Forehead-Nose** Mauchly's test showed a violation of sphericity against attention type ($0.24, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.53$) values. A large significant effect of attention type on the Forehead-Nose difference ($F_{1.6,34.7} = 29.08, p < .001, ges = 0.85$) was found. Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating attention, and between the selective and divided attention.

**Cheeks** Mauchly's test showed a violation of sphericity against attention type ($0.31, p < .05$), so we report Greenhouse-Geisser-corrected ($GGe = 0.57$) values. A large significant effect of attention type on the cheeks temperature increase ($F_{1.7,36.1} = 25.59, p < .001, ges = 0.64$) was found. Bonferroni-corrected post-hoc tests found a statistically significant difference between all attention types ($p < .05$), except between the sustained and the alternating, and between the alternating and the selective attention.

### *Effect of Visual task on Facial Temperature*

**Forehead-Nasal Temperature** we tested the effect of different attention type of visual tasks on the temperature metrics. Mauchly's test showed a violation of sphericity against attention type $(0.36, p < .05)$, so we report Greenhouse-Geisser-corrected $(GGe = 0.51)$ values. A large significant effect of attention type on the Forehead-Nose difference $(F_{1.9,38.9} = 50.5, p < .001, ges = 0.78)$ was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention.

**Cheeks** Mauchly's test showed a violation of sphericity against attention type $(0.30, p < .05)$, so we report Greenhouse-Geisser-corrected $(GGe = 0.58)$ values. A large significant effect of attention type on the cheeks temperature increase $(F_{1.8,36.7} = 30.05, p < .001, ges = 0.67)$ was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention, except between the sustained and the alternating, and between the alternating and the selective attention.

### *Effect of Audio-Visual task on Facial Temperature*

**Forehead-Nasal Temperature** Lastly, we tested the effect of different attention type of combination of audio-visual tasks on the temperature metrics. Mauchly's test showed a violation of sphericity against attention type $(0.39, p < .05)$, so we report Greenhouse-Geisser-corrected $(GGe = 0.39)$ values. A large significant effect of attention type on the Forehead-Nose difference $(F_{1.6,33.9} = 56.15, p < .001, ges = 0.88)$ was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention, except between the alternating and the selective attention.

**Cheeks Temperature** We tested the effect on the cheeks temperature with a one-way ANOVA. A large significant effect of attention type on the cheeks temperature increase $(F_{2.2,46.9} = 20.00, p < .001, ges = 0.59)$ was found. Bonferroni-corrected post-hoc tests found significant differences between all types of attention, except between the sustained and the alternating, and between the alternating and the selective attention.

In summary, our findings from the statistical analysis validate the correlation between attention types and cognitive load, deduced from the temperature changes in the selected region of interest. However, not all tasks exhibited significant difference between the alternating and selective attention types. Further, these findings are also aligned with the results from our subjective measure of perceived workload (NASA-TLX).

| Classifier | Classification Model | Gaze Features | Thermal Features | Gaze+Thermal Features |
|---|---|---|---|---|
| User Dependent | SVM | 57.2 ± 3.1% | 68.8 ± 3.8% | 52.4 ± 2.5% |
| *(Condition Independent)* | KNN | 58.4 ± 1.7% | 66.6 ± 3.4% | 52.2 ± 1.3% |
| | Logistic Regression | **58.9** ± **2.0%** | **70.7** ± **2.3%** | **77.4** ± **2.6%** |
| User Independent | SVM | 59.6 ± 2.2% | 70.6 ± 2.7% | 61.1 ± 1.8% |
| *(Condition Dependent)* | KNN | 61.3 ± 2.4% | 71.0 ± 3.7% | 61.3 ± 2.4% |
| | Logistic Regression | **76.5** ± **2.2%** | **69.7** ± **3.1%** | **86.9** ± **1.8%** |
| User Independent | SVM | 53.1 ± 1.9% | 72.3 ± 4.1% | 54.5 ± 2.5% |
| *(Condition Independent)* | KNN | 54.1 ± 2.1% | 72.5 ± 3.4% | 59.9 ± 2.2% |
| | Logistic Regression | **56.9** ± **0.9%** | **72.7** ± **2.6%** | **75.7** ± **1.8%** |

**Table 5.5:** Classification results.

## Classification Performance

To measure the performance of the classifiers, we computed the accuracy and Area Under the Curve (AUC), which aggregates precision and recall into one metric. We investigated the effect of the features used (gaze-only, thermal-only and gaze+thermal) as well as the usage of user-dependent and user-independent classifiers (condition-dependent and condition-independent) on the classification of attention types.

### *Comparison of Different Classification Models*

We first compared the performance of the classifiers for the attention types on the three different models: SVM, KNN and Logistic Regression. Table 5.5 shows the performance of the user-dependent and user-independent classifiers using the AUC score for the three classification models. The AUC score reported in the table is the average AUC for all the four task. As shown, overall the Logistic Regression model outperforms both SVM and KNN for all three feature sets (gaze-only, thermal-only and gaze+thermal). The reason being that KNN is an example of a lazy learner [18] classifier which memorizes the training data rather than learning discriminative function and its performance is highly dependent on the selection of k values passed as an input parameter [97]. Similarly, SVM classification results highly depend on the kernel and hyper parameters chosen. As for the Logistic Regression model, it has less generalization error than KNN and is easier to build compared to an SVM model [64], and for our purpose, it gives the best classification performance overall. Due to this reason, for the remainder of our analysis we have chosen to explore our results using the Logistic Regression Classification model.

| | | Gaze Features | | Thermal Features | | Gaze+Thermal Features | |
|---|---|---|---|---|---|---|---|
| Classifier | Task | Accuracy | AUC | Accuracy | AUC | Accuracy | AUC |
| User | Stroop | 44.3 ± 2.8% | 62.8 ± 1.5% | 57.9 ± 3.5% | 71.9 ± 3.1% | 68.3 ± 3.1% | 78.8 ± 3.3% |
| Dependent | Audio | 25.2 ± 1.8% | 51.1 ± 1.0% | 57.3 ± 3.1% | 71.4 ± 3.3% | 58.2 ± 1.2% | 71.8 ± 1.2% |
| (Condition | Visual | 46.2 ± 2.9% | 52.1 ± 2.8% | 54.8 ± 2.5% | 68.1 ± 1.3% | 70.4 ± 3.5% | 80.3 ± 2.3% |
| Independent) | Audio Visual | 54.4 ± 4.2% | 69.7 ± 2.8% | 57.9 ± 1.2% | 71.4 ± 1.3% | 68.8 ± 2.7% | 78.8 ± 3.8% |
| User | Stroop | 63.6 ± 4.6% | 75.8 ± 3.0% | 54.5 ± 4.5% | 69.7 ± 3.0% | 81.8 ± 3.5% | 87.9 ± 2.2% |
| Independent | Audio | 26.1 ± 1.1% | 50.8 ± 0.8% | 53.4 ± 4.7% | 69.7 ± 3.0% | 48.9 ± 3.0% | 65.9 ± 2.0% |
| (Condition | Visual | 78.4 ± 4.1% | 85.6 ± 2.8% | 54.4 ± 2.4% | 69.7 ± 3.8% | 95.5 ± 2.7% | 97.0 ± 1.8% |
| Dependent) | Audio Visual | 90.9 ± 3.1% | 93.9 ± 2.1% | 54.5 ± 4.2% | 69.7 ± 2.8% | 95.5 ± 2.1% | 97.0 ± 1.4% |
| User | Stroop | 45.5 ± 4.1% | 63.6 ± 1.0% | 59.1 ± 3.2% | 72.7 ±2.1% | 67.8 ± 1.6% | 78.8 ± 1.1% |
| Independent | Audio | 25.0 ± 0.0% | 50.0 ± 0.0% | 59.1 ± 5.5% | 72.7 ± 3.7% | 54.1 ± 2.7% | 69.4 ± 1.8% |
| (Condition | Visual | 40.1 ± 1.7% | 58.7 ± 1.1% | 59.1 ± 4.2% | 72.7 ± 3.3% | 70.1 ± 2.8% | 80.4 ± 1.0% |
| Independent) | Audio Visual | 36.0 ± 1.7% | 55.3 ± 1.4% | 58.2 ± 5.0% | 72.7 ± 1.3% | 61.8 ± 2.8% | 74.0 ± 3.1% |

**Table 5.6:** Logistic regression classification performance (all Stimuli).

### *Comparison of Different Classifiers (using Logistic Regression)*

Table 5.6 shows the overall performance for classification according to tasks for all classifiers. Overall, the user-independent, condition-dependent classifier performs the best compared to the other two classifiers with an average AUC score of 86.9%. In practice, this would be a classifier that is built-into the application, working for only for that application, but for any user. The high performance in this condition is expected due to the fact that this classifier is trained and evaluated on the same condition hence giving a higher performance for the same condition but not necessarily generalizing to other conditions. To build a more generalized classifier we built two other classifiers which are independent of the condition — user-dependent (condition independent) and user-independent (condition independent). We found the performance results to be comparable, obtained an average AUC score of 77.4% and 75.7% respectively. We note that these scores only decreased slightly when compared to the condition-dependent classifier, suggesting the validity of the general approach of using gaze and thermal imaging for attention classification. The user-dependent (condition-dependent) classifier is expected to perform slightly better as it will be trained and evaluated on the same user in the same condition. However, in the context of this work, we did not have enough data to train such a classifier.

Further, the results show that the accuracy of thermal-based classifier remained almost the same across all tasks. This means that the performance of the thermal-based classifier is largely independent of the task being performed by the user. Moreover, our findings showed that sustained attention required the

| Classifier | Task | Gaze Features | | Thermal Features | | Gaze+Thermal Features | |
|---|---|---|---|---|---|---|---|
| | | Accuracy | AUC | Accuracy | AUC | Accuracy | AUC |
| User Independent (Condition Independent) | Stroop | 46.5 ± 3.1% | 65.2 ± 2.7% | 48.9 ± 3.0% | 68.2 ±3.5% | 77.3 ± 4.0% | 84.8 ± 2.7% |
| | Visual | 44.1 ± 3.6% | 62.9 ± 2.4% | 59.1 ± 4.2% | 72.7 ± 3.3% | 79.5 ± 3.5% | 86.4 ± 2.0% |
| | Audio Visual | 37.5 ± 2.7% | 58.3 ± 1.8% | 57.9 ± 5.0% | 71.9 ± 3.3% | 62.5 ± 3.10% | 75.0 ± 2.3% |
| User Dependent (Condition Independent) | Stroop | 48.0 ± 2.5% | 68.3 ± 1.6% | 62.5 ± 3.2% | 75.0 ± 3.8% | 77.8 ± 3.6% | 85.1 ± 2.4% |
| | Visual | 48.9 ± 1.9% | 65.9 ± 1.3% | 60.3 ± 4.2% | 73.5 ± 3.5% | 80.7 ± 3.2% | 87.1 ± 2.1% |
| | Audio Visual | 57.9 ± 3.8% | 71.9 ± 2.5% | 68.1 ± 4.9% | 78.8 ± 3.3% | 76.1 ± 4.1% | 84.1 ± 1.7% |

**Table 5.7:** Logistic regression classifier performance without audio task.

least cognitive load followed by alternating, selective and divided attention, as reflected in the thermal features, and as suggested by our subjective measures. One important finding we observed was that the attention types are most accurately classified with an accuracy of (95.45%) when the participant is performing the visual and least accurately classified when the task being performed is audio. We also observed the same trend when comparing the performance of the user-independent (condition-dependent) classifier trained on just the gaze features. From our results, we observe that when classifying audio-only tasks, the thermal features alone worked as a better predictor than the classifiers that both gaze and thermal features. The reason being that the audio-only tasks lacks any visual stimuli, hence, the gaze features does not hold any significance for classifying attention in audio-only task—effectively working as noise. With reference to Figure 5.15, we can see that all tasks in each attention type have a distinct pattern, for example, in the alternating attention tasks, we can see a clear pattern the left and right AOIs. As for the audio-only tasks, the gaze patterns appear to be random with the participant, either focusing randomly around the screen or at a focused point with random saccades around the screen. Due to this reason, the average for the condition-independent classifiers for all tasks does not perform as well compared to the condition-dependent classifier as their training set includes the insignificant gaze features of the audio-only task, which decrease the classification accuracy for the gaze-only and gaze+thermal feature sets.

To measure the effect of removing the task which lacks visual stimuli (i.e. audio-only tasks) on the condition-independent and user-dependent classifier, we retrained our classifiers by only considering the task with visual stimuli (Audio Visual, Visual and Stroop). This is so the gaze features in the training data set would remain meaningful in the classification process.

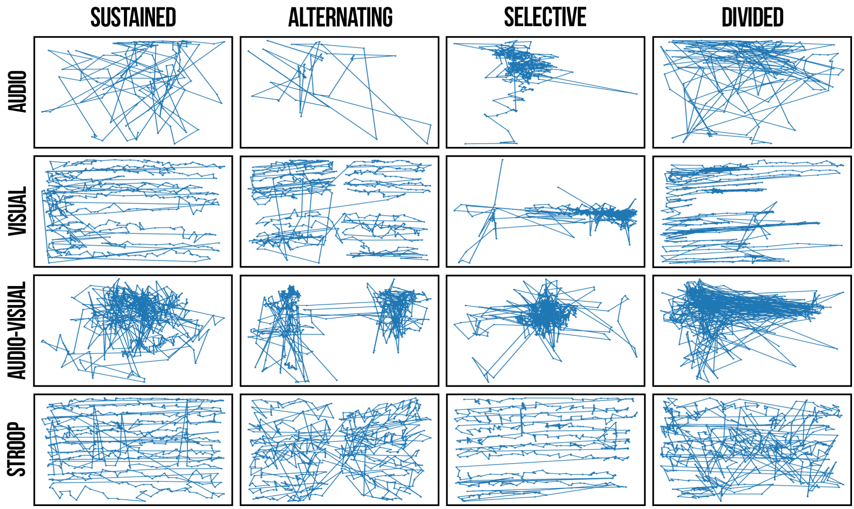



**Figure 5.15:** Gaze Plots, highlighting the patterns in each task.

We evaluated the *user-independent, condition-independent* classifier by training it 22 times using leave-one-out-cross-validation (LOOCV) three times, one for each condition. Each time, we trained it on the data of the 21 participants for two conditions and evaluated it on the data of the third condition from the last participant. As for the user-dependent, condition-independent, classification, we trained the classifier 22 times using cross-validation three times, one for each condition. Each time, we trained it on the data of the single participant for two conditions and evaluated it on the data of the third condition from the same participant. The results of the user-independent, condition-independent and the user-dependent, condition-independent classification that only considers tasks with visual stimuli shows an increased classification accuracy (see Table 5.7). Hence, observing a better classification accuracy for each task for both type of classifier has increased as now the gaze feature plays a significant role in attention classification. In practice, this means that eye-tracking only helps the classifier when the task involves a visual stimulus. Otherwise (i.e. as in audio-only tasks), it tends to harm the classification results.

## 5.2.4 Discussion and Summary

The results of our study and a review of prior work revealed that different attentive states could be distinguished by the locus of visual attention and estimated cognitive load. On this basis, in the previous sections, we presented the results from collecting, analyzing and classifying gaze and thermal data of different attention types, which we summarize and discuss grouped by the most important observations in the following.

### On Performance

In this work, we discuss a first attempt of combining thermal imaging and eye tracking to discriminate between four types of user attention. Our results show that attention classification is feasible, achieving an accuracy of up to 95.45% when using a condition-dependent prediction (see Table 5.6). This result is promising as it paves the way for new applications in which classification can be tailored to a particular known condition or task. For example, this could be embedded into an e-learning system to measure student attention during a lecture.

In contrast, the condition-independent classification is more challenging. When comparing the performance with the condition-dependent classifier, we observed a decrease in the accuracy between 62.5% and 79.50% while considering only the tasks with visuals stimuli. Though this performance might be sufficient for some applications and is well-above the 25% baseline, further work is needed to bring performance up to the same level as for condition-dependent prediction. This means that this approach is not yet quite feasible for distinguishing attention types in unknown tasks.

However, in all of our experiments, our user-independent results were strong, suggesting that by training the classifier on one specific task, the classification generalizes well to unseen users.

### On Discriminating Different Attention Types

Based on our review of the literature, we hypothesized that different attention types require different cognitive load levels, which would lead to a change in the participants' facial temperature patterns. From previous work, we know that regions on the face such as forehead, nose, and cheeks are often visible points

| Attention Types | Gaze | Thermal | Gaze & Thermal |
|---|---|---|---|
| Alternating | 100% | 27.27% | 100% |
| Sustained | 13.0% | 86.36% | 90.9% |
| Selective | 27.0% | 36.36% | 18.18% |
| Divided | 50.0% | 69.7% | 81.81% |

**Table 5.8:** Recognition accuracy of each attention type.

and are feasible for temperature measurement [13, 117] We tested the effects of different tasks and their attention types on temperature changes in these points. We elicited different attention types through set of tasks with different stimuli and found a significant difference in the metrics used across the different attention types. We confirmed the validity of our findings, where the same pattern of facial temperature changes was observed across the different conditions (Stroop, Audio, Visual, and Audio-Visual). Our findings from the statistical analysis validate the correlation of cognitive load, deduced from the selected region of interest, and the different attention types. Although it was not significant across all attention types (e.g. sustained and alternating attention), it could give a hint about the experienced attention type. Furthermore, this highlights the role of gaze data to complement thermal data.

For discriminating attention types using a classifier, we investigated the performance when classifying each attention type separately. Our results show that alternating attention achieved the highest accuracy for the thermal and the eye feature set because the alternating gaze pattern of participants from one AOI to another is a strong indicator of alternating attentional state. For this task, the thermal features do not capture much information of participant attention state as indicated by low performance of the classifier trained on just the thermal features set (see Table 5.8).

For sustained and divided attention, gaze features did not work as well but we found that thermal features performed well. Temperature variation was considerably different compared to other attentional states as shown in Figure 5.14. For selective attention, the performance of the classifier was the lowest. This attention type was mostly confused with divided attention, as can be seen from the confusion matrix (see Figure 5.16). One likely reason is that the thermal features (forehead-nasal and cheeks) change across all attention types. For instance, the change in facial temperature for selective attention overlaps the most with the divided attention (Figure 5.14).

## On Combining Thermal Imaging and Eye Tracking

Additionally, we found that combining both gaze and thermal features boosted the performance of the classifier as compared to using gaze or thermal only (Figure 5.16 and Table 5.7) for the visual tasks. This is because each modality complements the other for the classification of attentional types. For instance, Figure 5.15 shows that divided and sustained attention present very similar gaze patterns but elicit very different levels of cognitive load, which is reflected in the thermal features (Figure 5.14). In contrast, alternating attention presents itself somewhere in between sustained and selective attention in terms of facial temperature but exhibits very distinct eye movement patterns as reflected in the figure. This highlights the importance and potential of using thermal imaging and eye tracking in combination to classify attention types. Interestingly, we observed that thermal features exhibit the same performance for different conditions. This validates that different attention types allocate different cognitive load, regardless of the stimuli (see Table 5.6). In contrast to the gaze features, because we rely on the AOI, the features are influenced by the task and stimuli. This is reflected in classification accuracy using only the gaze features. For instance, as shown in Table 5.6 the fixations obtained in the audio task for various attention types were arbitrary (see Figure 5.15), and we did not observe any unique patterns of gaze transition for different attention type as the participant was asked to just attend to the playing audio.

## On Different Conditions

We observed a decrease in classification accuracy for the audio condition when using gaze and thermal as opposed to using thermal features only. This is because in the audio condition, participants' eye movements were arbitrary, due to the lack of visual stimuli. Hence, training the classifier with the audio task gaze data, would mean training the classifier with confusing data. In other words, including gaze data of the audio only condition, would then yield to reduced performance. Further exploring the confusion matrix (see Figure 5.18) of the classifiers trained on the thermal feature only, we can conclude that for an audio-only task attention could be classified into sustained and divided attention more accurately then the selective and alternating type. Based on this observation, we suggest using only thermal features to classify attention types for audio-only tasks.

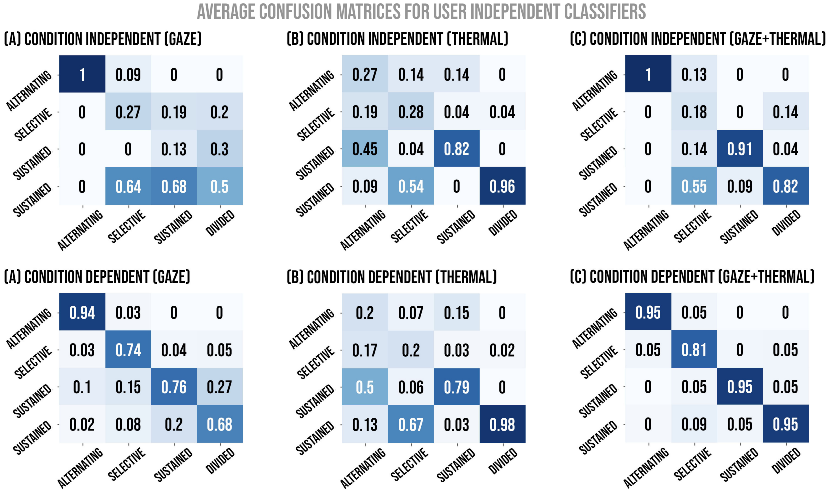



**Figure 5.16:** User Independent Average Confusion Matrices.

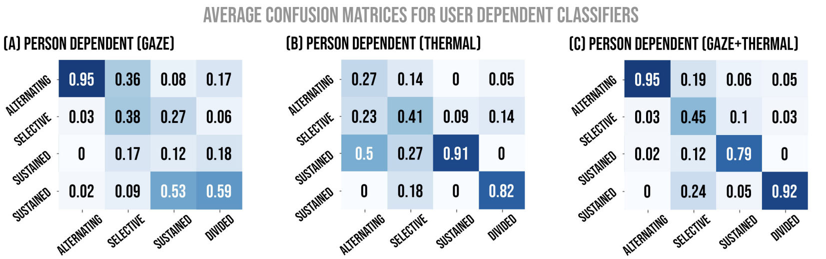



**Figure 5.17:** User Dependent Average Confusion Matrices.

## On Attention Type-Aware System Development

We observed that the average classification accuracy for the user-dependent, condition independent classifier in all four tasks and three feature sets is higher than the user-independent classifiers (see Table 5.6 and Table 5.7). This means that the user-dependent classifier was able to predict the attention type of a specific user more accurately when trained on the data of the same user (user-dependent) rather than training it on features of all users (user-independent). Similar results on discriminating attention types were achieved for a user-dependent classifier

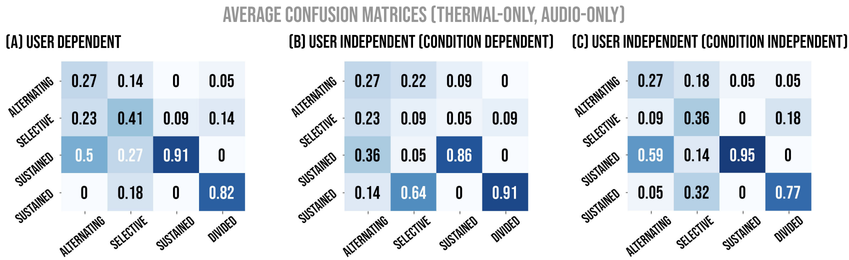



**Figure 5.18:** Average confusion matrices for all classifiers for audio only task.

(see Figure 5.17) with alternating attention achieving the highest accuracy, and selective attention achieving the lowest accuracy for thermal and gaze feature set. We found that similar to the results obtained for user-independent classifier, the performance of user-dependent classifier is also boosted by combining both gaze and thermal feature set for the Stroop, Visual and Audio-Visual task. One obvious limitation of user-dependent classifier would be that it would not be generalized for different users. In practical terms, this means that if the system requires user calibration prior to use. However, if a real-time attention classification system is required, which could classify attention of any user without being trained every time for a new user, then a user-independent classification approach would be more suited. Therefore, the types of classification approach taken would highly dependent on the type of application the system is used in.

## Example Use Cases

Our findings show that the proposed classifier was able to classify attention types unobtrusively. Applications that take into account the attention type can be applied to a broad range of applications ranging from education [102, 212], performance management in the workplace, distraction management [27, 139], to quantified-self applications [62]. Educational applications could monitor students' attention type and adapt accordingly, e.g. assessing if the presented material is “attention-grabbing" so that the students would show sustained attention rather than divided or alternating attention, aiming to better design learning systems [102]. Furthermore, workplaces could benefit from our approach by helping workers manage their attention if he/she are experiencing divided attention during safety-critical task, and to avoid divided attention in dangerous situations (e.g. operating trains) [156, 284]. Additionally, if a user should focus to finish a task, an attentive user interface could help the user to keep their attention sustained on the task [234].

Online distractions are a controversial aspect of our current technology-mediated workplaces. Our approach could be used to manage distractions, especially with users that are more susceptible to social distraction [167]. An interface could block the distraction source (e.g. social media, smartphone) when alternating or divided attention between the task in hand and the distraction source is detected [27]. We also can deploy attention-type detection for quantified-self applications as proposed by Dingler et al. [62]. The system could monitor the attention type patterns throughout the day, aiming to assist users with tracking and managing their attention distribution to enhance their well-being. For instance, stress and frustration occur when there is a mismatch between the accomplished tasks and the planned ones [244] due to the lack of sustained attention on the planned tasks. Furthermore, the high frequency of divided attention may lead to burnout and memory distortion [177].

## Limitations and Future Work

This work proposes the first steps towards classifying attention using unobtrusive sensors. As such, we required a dataset with clearly labeled attention types for training our classifier.

Despite these promising results, our work has several limitations that we plan to address in future work. First, the controlled task is likely to lead to behavior changes. Similar to studies in Affective Computing, there is a trade-off between the quality of the labels and the naturalness of user behavior. We opted for a controlled setup to increase the quality of the labels at the expense of natural behavior. By demonstrating the feasibility of the approach, our next steps will involve collecting a more naturalistic *in-the-wild* dataset. Second, we labeled the data according to the elicited attentional state. While these tasks were informed by previous work in psychology, it is difficult to guarantee that users were in those states at all times during the tasks. For example, during sustained attention, we cannot guarantee that participants did not momentarily “mind-wandered”. We tried to minimize these effects by keeping our tasks time reasonably short. Third, users’ eye movements are highly dependent on the stimuli used. We attempted to minimize the effects of the particular stimulus by abstracting from the visual layout of the interface , instead, computing features based on Areas of Interest (AOI). This tends to minimize the overfitting due to the visual design of the interface as compared to low-level features only. For example, in our design, the two pieces of text in the alternating attention condition were side-by-side. If we had trained a classifier using saccade directions, a high proportion of large sideways saccades would likely be indicative of alternating attention. However, if the same classifier were applied in an interface where the two texts were displayed one on top of each other, the approach would no longer work. Using AOIs allows us to abstract from the specifics of the interface, but also introduces a new challenge—how to determine which areas are of interest. This limitation can be addressed in many ways. For example, a learning system could specify that the video player is an AOI. A system like *RescueTime* could classify the applications that are part of productive (i.e., attentive) use of time and set it as the AOIs.

Fourth, thermal imaging is influenced by external factors, e.g. changes in room temperature, and internal factors, e.g. changes in affective states. These can be confounds that might affect the performance of the system in the wild. A more naturalistic dataset is required to explore these questions. Additionally, we envision that running an evaluation on participants with more experience in executing focused tasks such as seasoned workers would yield interesting insights, as well as running this over a longer period of time. We also plan to explore the performance of the gaze and thermal classifier by extracting more stimulus-dependent gaze features such as saccade velocity and length from one AOI to another and stimuli dependant feature such as the total fixation rate and mean saccade rate and angle for the individual task. Lastly, our findings open up further research question—how to distinguish between selective and divided attention. This can be explored by augmenting another bio-data e.g., GSR, heart rate, aiming to investigate if they differ in terms of other physiological responses.

## Summary

Through our review of related work, we concluded that no prior work explored the use of thermal imaging combined with eye tracking to classify attention types. Consequently, in this work, we began our exploration by identifying gaze and features that could potentially reveal the four attention types—*sustained*, *selective*, *alternating* and *divided* attention. We investigated the effects of using different feature sets (gaze, thermal and the combination of thermal and gaze features) in classifying the four attention types. We used the extracted features to train two categories of classifiers: (1) condition-dependent and (2) condition-independent classifiers. Our classifiers achieved AUC up to 95.45% and 79.5% respectively. Furthermore, we investigated the performance of user dependent and independent classifiers, we had AUC up to 75.7% for the user independent-condition independent, 87% user-independent-condition dependent, and 77.4% for the user-dependent classifier). We additionally found that there is an increase in the classification accuracy when using the combination of gaze and thermal features as opposed to using gaze or thermal features alone. In this work, we were able to classify attention types unobtrusively, using a thermal camera and a remote eye tracker. This enables novel opportunities in the field of attention-aware computing: our approach, for example, can be applied in different research areas, e.g., education, adaptive and assistive systems. It could also be used to track and give feedback to the user, to increase the user's awareness of their attention patterns.

In summary, our results reveal the feasibility of building an attention classifier based on facial temperature and eye movements. Hence, we envision that our work can serve as an initial building block to understanding the human mind and the influence of different attention types. We hope that developers of attention-aware and adaptive systems can use our results to build enhanced adaptive systems with a diverse set of application to benefit users in everyday usage.

# III

# Design Implications & Guidelines

# Chapter 6

# Implications and Design Recommendations

This thesis contains ten in-depth research probes informing the design of systems using thermal imaging, particularity to amplify perception. In this chapter, we present design recommendations distilled from the previous chapters, to address both social and technological perspectives to support using thermal cameras in diverse setups, and to inform researchers and designers. We derive a set of implications for designing future thermal camera based systems and tools that will consider the challenges and recommendations identified in our findings. In addition to the individual recommendations in each of the presented Chapters, we highlight the following recommendations for designing thermal imaging.

## 6.1 Introductory Phase

Amplified perception and generally the usage of thermal cameras are influenced by knowledge bias, in that users must have an initial understanding and knowledge to best utilize the amplified perception. In our exploration participants intuitively explored the different perception mode to have an impression of its capabilities. Future design should support a "learning/introductory phase" prior to usage or during setup to ensure full understanding of the alternative amplified mode.

Further challenges lie in the image and information representation. Presenting raw image data might not be intuitive for users, therefore, designing such tools should consider processing the raw image data and try to present the users with more meaningful information. The thermal images are considered to be straight forward.

## 6.2 Application Specific Form Factor

Systems should use the appropriate form factor for the use case for which they are designing their system. Where it is a monitoring use case, participants preferred the mounted stationary form e.g. for continuous health monitoring or object state detection. On the other hand, our findings also recommended the mobile form factor for on-demand explicit usage of vision extension and thermal cameras e.g. checking a baby.

## 6.3 Context Awareness and Social Context

Designers should consider the context of use and offer information related to the context (e.g. highlighting the hottest cup, or automatically detecting and displaying someone's emotional state). This context awareness would enhance the understanding and usage of thermal camera based systems, as it would help novice users learn about the capabilities of the camera and hence be able to best utilize it. Our findings suggest the need of including an information layer in the thermal user interface.

## 6.4 Privacy and Social Consideration

Participants exhibited an awareness of the privacy implications of having such a layer of extra information at hand. This emphasizes the need to explore and research explicit privacy management in HCI for thermal cameras or any imaging system that displays non-visible information. For instance, designers should consider camera state notification e.g. an indicator when the camera is on.

Additionally, users displayed a high level of awareness on the social aspects and privacy implications of having such a layer of extra information at hand and commented on how it could be a potential means of discrimination (for example detecting and avoiding a peer with a fever, based on their temperature). This opens an HCI challenge in designing such tools, as privacy and ethical considerations are raised by using such a tool.

Naturally, there are many ethical concerns when considering collecting and viewing data about humans' normally don not perceive. As the captured information could be revealing internal and private information e.g., smartphone PIN or affect state. Thus, protecting the privacy of individuals is a major concern and important field of investigation when utilizing thermal imaging as a sensing technology. The minimal amount of data needed for extracting the necessary information should be determined for each use-case. Existing measures for protecting the individuals and maintaining their privacy, should be applied. For instance, to maintain privacy in the context of building adaptive systems based on the facial temperature, one possibility is to analyze the data in the client side of applications and thus storing as little information as possible about the individuals in global databases. Depending on the applications and the accuracy of the models representing it, the amount of data to be processed at the client side can be determined. While there are privacy concerns associated with collecting affect data in general, additional measures need to be investigated for thermal imaging. Existing solutions can be be employed, but should be extended given the needs of the current use-case and should account for the non-visible spectrum.

## 6.5 Ethics and Data Collection

Although utilizing thermal imaging in building novel systems is appealing, one may not forget that its commercial application may not come without ethical concerns. Starting from the data collection, as presented in Chapter 5, thermal imaging could act as a window into our souls and mind, by inferring our internal states based on our facial temperature. Hence, if privacy and ethical consideration are not guaranteed to individuals then it is very unlikely that many participants would be willing to take part in any experiment. Thus, we need to guarantee full anonymity and data privacy maintenance to encourage individuals to provide us with the needed data.

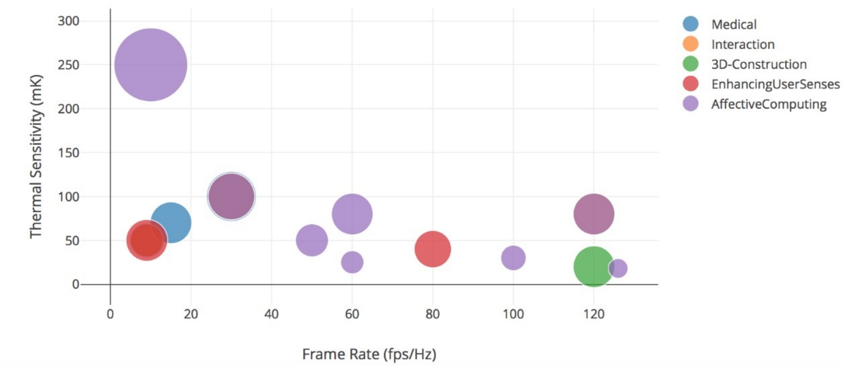



**Figure 6.1:** Thermal camera technical specifications for divers applications.

## 6.6 Technical and Generic Conceptual Architecture

In this thesis we built the study probes using different models of thermal cameras. Each thermal camera has different technical specification in terms of thermal sensitivity as well as frame rates. Figure 6.1 presents these values from the used thermal cameras in the previous chapters. Further, we conducted an intensive literature review, to distill the technical specifications of thermal cameras deployed in various research fields. Our review focused on the research conducted under the umbrella of HCI, to identify the technical requirements and specifications of thermal imaging. We exclusively used ACM and IEEE digital libraries as our source, using the keywords *Thermal Imaging, FIR and HCI*. Another review and classification iteration was conducted on the cited papers in each retrieved paper to ensure that we covered the related research conducted over the past twenty years. Taking a deeper look into thermal imaging-related publications, we find that these technical specifications vary considerably in the application domains. For instance, the 3D reconstruction domain requires a high frame rate yet the thermal sensitivity does not play a significant role. Interestingly, the affective computing domain deployed thermal cameras with varying thermal sensitivity but most used high thermal sensitivity and frame rate. While we aimed to derive technical guidelines for developers and researchers, most of the conducted research used the high end thermal imaging as the commercial options have been available only recently. Researchers should consider the values and ensure that their proposed application is feasible with the acquired hardware.

**Generic Conceptual Architecture** In all the built study probes, we followed the same processing pipeline as listed below, as well as the usage of open source libraries e.g. OpenCV. The main aim behind this approach is to provide a generic architecture for thermal imaging-based systems. Providing an easy to use software is one of the key aspects to encourage the development of systems. The system should provide possibilities for beginners to rapidly develop small applications with low complexity but should allow a low level access to the raw data. Hence we propose the following generic conceptual thermal imaging processing pipeline:

1. **Frame extraction:** Frame extraction is the initial step to sample and extract the frame from the camera feed, based on the camera's frequency.
2. **Image pre-processing :** We apply the supported image pre-processing techniques from OpenCV, to enhance the quality of the captured thermal frames for further processing.
3. **Feature Extraction & Recognition:** In this step, features of interest are extracted using OpenCV. The features are to be determined by the designers and developers, for instance the hottest object in the scene or the facial temperature of the user.
4. **Data Presentation:** Designers and Developers have full control over the data presentation. This dimension should be application specific. In particular, implicit thermal data presentation should be applied when users are intended to visualize the heat map e.g. detect the touch traces. On the other hand, explicit data presentation is required when the thermal features have to be interpreted before being presented to the users e.g. cognitive load level.
5. **Temperature and Raw Data Recording:** Lastly, developers and researchers should consider recording the captured data for off line analysis as well as the creation of novel datasets.

# IV

# Conclusion and Outlook

# Chapter 7

# Conclusion and Outlook

*This chapter is based on the following publication:*

- Y. Abdelrahman and A. Schmidt. Beyond the visible: sensing with thermal imaging. *Interactions*, 26(1):76–78, 2019

This thesis explored the usage of thermal imaging to amplify human perception of the environment as well as the states of others. Whether through HMDs, stationary setups, or mobile form factor, we investigated the various opportunists and challenges of perceiving the world in the thermal spectrum. We followed the user-centered design process extended through a probe-based approach for understanding the usage of thermal imaging, requirements and constraints, aiming to derive system and design requirements. In this chapter we summarize our research contributions, and provide a conclusion and future outlook for thermal imaging-based interactive systems that enables users to amplify their perception.

## 7.1 Summary of Contribution

Overall, this thesis provides five main contributions. First, we apply and extend the user-centered design process to explore the opportunities and challenges of designing amplified perception using thermal imaging. Second, we present study probes and prototypes increasing the understanding of the opportunities of thermal imaging and how they should be deployed. Third, we provide a set of design recommendations helping to design and deploy thermal imaging into interactive systems. We present a summary of our findings in light of our research questions on the user and technical levels, synthesized through our findings. Fourth, we present a reference recommendation for technical decisions. Additionally, we present a conceptual generic platform architecture that aims to help developers of thermal imaging-based applications, to gain a better understanding of the requirements and design aspects they need to consider. Finally, we release the generic algorithm for thermal imaging analysis, as well as the collected dataset throughout the conducted research in this thesis, for future researchers and the community to utilize and build upon.

### 7.1.1 Understanding Thermal Imaging

We applied the user-centered design process to the field of thermal imaging. We investigated and analyzed how diverse users understand and perceive the thermal spectrum to highlight the opportunities and challenges of thermal imaging, as well as to identify the capabilities of amplifying perception. In a first step, we highlighted the specific context of use and its characteristics. To specify the requirements from a technical and user point of view, we extended the user-centered design process with a probe-based research approach. By designing, implementing, and evaluating research probes tackling different amplified perception possibilities, we learned fundamental aspects that resulted in design recommendations.

We undertook a research probe approach, where a prototype was developed to evaluate a certain opportunity. We presented ten probes; six probes investigating the amplified environmental perception, along with two probes exploring amplifying our perception of cognitive load.

### 7.1.2 Amplified Perception of the Environment

Effective user understanding of the thermal spectrum was the focus of our first three presented probes. We looked into presenting a thermal view using HMDs and on demand mobile application. While both means presented thermal information back to the users, we learned different lessons from each evaluation. From the HMDs, we found that users appreciate the hands-free amplified perception and envisioned a futuristic form factor for amplified perception. However, from the mobile probe we discovered that users prefer having it on demand with an explicit action to view the environment in an amplified form.

The developed research probes provide insights into how the perception and awareness can be enriched with thermal imaging. Nowadays tools enhancing our visual perception rely on the visible light enhancement. In contrast, we outline possibilities currently not explored using thermal imaging. While technical challenges with regards to miniaturization and robustness still need to be tackled, we show the generalfeasibility of the amplified perception.

### 7.1.3 Amplified Perception of Cognitive Load

Including the hidden internal states in our perception was explored through two probes. In one probe we explored the usage of thermal imaging to reveal the cognitive load level. In the second probe we complemented current technologies namely eye trackers to reveal the experienced attention type. The two probes showing the utilizing thermal imaging to capture and interpret our facial temperature outperformed existing technologies. Using thermal imaging was shown to be unobtrusive and operated in almost real time, giving it an edge over existing technologies to reveal insights about others' internal hidden states.

### 7.1.4 Design Implications and Guidelines

Grouping and analyzing the outcomes from our study probes and prototypes, we charted a set of design implications and guidelines concerning both user and technical perspectives. The outcomes of our research probes paved the path towards a deep understanding of the technology in hand and uncovered the set of design recommendations and conceptual generic architecture for thermal imaging based systems. Finally, we contributed a set of design recommendations derived from our findings and evaluations of all the study and prototype probes. The recommendations were then used to structure a conceptual architecture that can inform researchers, designers, and developers interested in the usage of thermal imaging to build novel interactive systems.

In summary, the contributions of this thesis could be classified into two main strands: 1) empirical, and 2) technical contribution.

## Empirical Contribution

The empirical contribution of this thesis contributes a body of knowledge that augments our understanding of thermal imaging and amplified perception. First, our work has identified opportunities brought forth by utilizing thermal imaging. Second, we explore core challenges that are unique to thermal imaging based systems. Accordingly, this thesis proposes systems that leverage said opportunities, and systems and studies that address and deepen our understanding of the challenges. At the same time, this work lays a foundation on which future studies and systems can build on: 1) Many of the identified opportunities are waiting to be explored and leveraged by researchers and practitioners. 2) Our exploration of unique aspects of thermal imaging has set the scene for the upcoming research challenges for integrating thermal imaging in novel interactive systems.

## Technical Contribution

We have designed various systems that leverage thermal imaging to build novel applications or to enhance existing ones. In this thesis, We provide a technical contribution by showing how a generic thermal image processing algorithm could be realized. We developed and released a set of prototypical open source systems e.g. *ThermalAnalyzer*, *ThermalMirror*, *CognitiveHeat*, *AttentCam*, *VID*, along with the corresponding dataset. The applications consist of a C# Windows Forms Application service and an openCV. The service handles the connection to the Optris thermal camera irrespective of its model, extracts areas of interest, interprets the data and provides visualizations of the thermal data for the developers.

## 7.2 Future Work

This thesis provides a common ground for future research in the area of thermal imaging based interactive systems. However, during the course of this thesis, several additional research challenges have arisen which could provide the basis of future research. This chapter explains these research challenges in detail and highlights the main directions suggested for future work.

### Long Term Studies of Thermal Imaging Usage

Through our work, we evaluated the temporary usage of thermal imaging as a perception amplifying tool. The studies were all limited to two weeks at most. Long term usage and investigations would strengthen our findings and help uncover technical and social effects of having amplified perception via thermal imaging on the private and public level. Furthermore, they would yield deeper insights into the usage behavior as well as the users' understanding of the thermal spectrum, revealing future opportunities. Users of our probes envisioned the usage of thermal imaging on a daily basis. However, social and privacy concerns were raised. Exploring social acceptability of viewing scenes in the thermal spectrum uncovering non-visible information in different contexts, environments, and spaces in long term studies would lead to a better understanding of the social and usage aspects.

### Exploring Thermal Images Visualizations

In our work, we explored only the color mapping of the thermal feed as the visualization technique. Additionally, we evaluated different visualization techniques, however these were limited to the firefighters perception, and included sensor fusion data from the depth and RGB cameras. A systematic and in-depth investigation of the overlay and different visualizations of the thermal data is complex and an opportunity for future work.

## Thermal Imaging and States Co-founds

Our work highlighted the ability of thermal imaging to unobtrusively estimate and classify affective states. However, systems using thermal imaging for affective or cognitive computing always include a factor of uncertainty: On the one hand, due to their dependence on other external factors (e.g. room temperature, time of day, hours of sleep, stress, food intake etc.), as well as internal factors, e.g. changes in other affective or emotional states. These can lead to have complications such as stress and cognitive load, and might affect the performance of the system, and might affect the accuracy of the system outcomes and introduce uncertainties. However, most of the systems aim to give a glimpse into users' affect states rather than exact accuracy. We believe that investigating the accuracy and uncertainty is a great opportunity for future work. This would lead to the development of more robust, state deterministic systems to enhance the perception of others' internal states to achieve the goal of revealing hidden inner states efficiently and unobtrusively.

## Validation of The Estimated Users' State

One challenge we have been faced with during our explorations of thermal imaging for amplified perception of others internal states, especially with non trained users, is validating the classified user data. Researchers proposed the usage of subjective measures and ratings (e.g. using existing questionnaires from psychology) to validate the classified state. However, this compromises the unobtrusiveness and real time operation offered by thermal imaging. Furthermore, not all classified states have precise questionnaires that aim to validate them, for example user attention type. While in our work and generally in HCI we often utilize these questionnaires from Psychology, it is currently becoming more and more inevitable that HCI researchers must develop their own standardized ways of validating information. This is a much needed opportunity for future work, for the gap is currently widening with new forms of sensing and inferring users' states, especially that it operates in an unobtrusive, remote and real time. Our work highlights the need of the tight collaboration of HCI and Psychology to design methods for validating sensed and interpreted information that is suitable for the current status of affective computing and HCI research.

## Evaluation In The Wild

Despite the promising results of the ability of thermal imaging to unobtrusively estimate and classify affective states, it opens the room for future work to address the raised challenges and limitations. Current evaluation methods are mainly limited to feasibility studies in the lab. While these evaluations include a high internal validity and show the general feasibility of the envisioned approaches, moving from the lab to field is the necessary next step. Additionally, the controlled task used is likely to lead to behavior changes. Similar to studies in Affective Computing, there is a trade-off between the quality of the labels and the naturalness of user behavior. In our work, we opted for a controlled setup to increase the quality of the labels at the expense of natural behavior. By demonstrating the feasibility of the approaches, future steps will involve collecting a more naturalistic *in-the-wild* dataset. Second, we labeled the data according to the elicited states. While these tasks were informed by previous work in Psychology, it is unclear how effective this elicitation was. These can influence the performance of the built systems in the wild. A more naturalistic *in the wild* evaluation is required to explore these questions and to investigate the ecologic validity.

## Future of Thermal Imaging

Zooming out to the bigger picture, over the past 10 years thermal imaging has penetrated the commercial market by its reduced price and size. We envision that thermal imaging will take even further leaps in the next 20 years, not only by the availability of the technology, but rather by the deep understanding of the users' perception of thermal imaging as well as the novel opportunities raised by the technology. In this thesis we explored how thermal imaging could be deployed to amplify our perception by revealing the temperature information of our vicinity. However, thermal imaging offers enhanced and novel opportunities in vast and diverse research fields. In our exploration we mostly used the Optris thermal camera. Although this is considered as a commercial thermal camera, yet it is a high end one. This might impact upon its feasibility and affordability. We envision that in the near future, the low end thermal cameras would have the same capabilities as the high end relatively more expensive commercial thermal cameras.

However, we believe that the more prominent challenge is not the feasibility of acquiring the technology, but rather how we deal with this new available information, what it reveals about us, and how it affects the way we see our environment, ourselves, and others. On one hand, thermal Imaging shows high potential in building novel systems, exploiting the thermal spectrum and making the invisible visible in a remote and contact-less manner. Yet, along with the opportunities it raises challenges and risks as well. We can imagine a future where technology misuse is exercised by revealing private information like the affect state of bystanders without their consent, or even making people more prone to authentication attacks, as generally, cameras have been widely known for their high level of social intrusiveness. This raises the challenges of data ownership and the lack of power of choice of self representation.

On the other hand, one can envision a more optimistic future, where the usage of such technology is governed and the positives and negatives of such a technology come into play. We believe that in 10 years a balance can be reached between the potential negative effects and the whole opportunities of such technology to change the way we perceive and interact with the world. Addressing the data ownership and privacy concerns, advanced algorithms would allow for data protection and ensuring that full control would be given to the user.

We believe that our exploration is merely the tip of the iceberg in a whole new wave of thermal imaging applications to hit the research field. Amplifying our visual perception is only one step towards utilizing thermal imaging for more contextual sensing and interpretation of our environment via thermal information, by empowering users to extend their perception.

## Concluding Remarks

This thesis investigates how thermal imaging can be used to enrich and amplify our perception concerning both the environment and our hidden internal states, in particular cognitive load. It addresses fundamental challenges designers and developers of such systems will face in the future, when we reach the point where thermal imaging is integrated in our daily devices e.g. laptop web-cameras and smart-phone integrated cameras for about the same price as the currently available devices. We saw the first steps made when CAT released a smart-phone with built in thermal cameras. However, development has not progressed far enough to penetrate the market and reach a competitive market sector with existing technology. We envision that eventually, research and developers especially with the vast move of the DIY, will be able to meet market requirements with regard to price and quality. From then on, understanding how this development influences the interaction and design of thermal imaging based systems becomes crucial.

# V

# BIBLIOGRAPHY

# BIBLIOGRAPHY

[1] Y. Abdelrahman, M. Hassib, M. G. Marquez, M. Funk, and A. Schmidt. Implicit engagement detection for interactive museums using brain-computer interfaces. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 838–845. ACM, 2015.

[2] Y. Abdelrahman, M. Hassib, M. G. Marquez, M. Funk, and A. Schmidt. Implicit engagement detection for interactive museums using brain-computer interfaces. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '15, pages 838–845, New York, NY, USA, 2015. ACM.

[3] Y. Abdelrahman, M. Khamis, S. Schneegass, and F. Alt. Stay cool! understanding thermal attacks on mobile-based user authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3751–3763. ACM, 2017.

[4] Y. Abdelrahman, A. Khan, J. Newn, E. Velloso, S. Safwat, J. Bailey, A. Bulling, F. Vetere, and A. Schmidt. Classifying attention types with thermal imaging and eye tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technology*, Sept. 2019.

[5] Y. Abdelrahman, P. Knierim, P. wozniak, N. Henze, and A. Schmidt. Designing with firefighters *CHI'19*.

[6] Y. Abdelrahman, P. Knierim, P. W. Wozniak, N. Henze, and A. Schmidt. See through the fire: Evaluating the augmentation of visual perception of firefighters using depth and thermal cameras. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous*

*Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 693–696, New York, NY, USA, 2017. ACM.

[7] Y. Abdelrahman, T. Kubitza, K. Wolf, N. Pohl, and A. Schmidt. Tick that box: Interactive paper documents. In *Human-Computer Interaction*, pages 376–383. Springer, 2015.

[8] Y. Abdelrahman, A. Sahami Shirazi, N. Henze, and A. Schmidt. Investigation of material properties for thermal imaging-based interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 15–18, New York, NY, USA, 2015. ACM.

[9] Y. Abdelrahman and A. Schmidt. The history of thermal imaging in human computer interaction *TOCHI*.

[10] Y. Abdelrahman and A. Schmidt. Beyond the visible: sensing with thermal imaging. *Interactions*, 26(1):76–78, 2019.

[11] Y. Abdelrahman, A. Schmidt, and P. Knierim. Snake view: Exploring thermal imaging as a vision extender in mountains. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 1067–1071, New York, NY, USA, 2017. ACM.

[12] Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere. Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technology*, 1(3):33:1–33:20, Sept. 2017.

[13] Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere. Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3):33:1–33:20, Sept. 2017.

[14] Y. Abdelrahman, P. wozniak, N. Henze, and A. Schmidt. Exploring the potential of thermal imaging usage at home by novice users *CHI'19*.

[15] Y. Abdelrahman, P. Wozniak, P. Knierim, N. Henze, and A. Schmidt. Exploration of alternative vision modes using depth and thermal cameras. In *Proceedings of the 17th International Conference on Mobile and*

*Ubiquitous Multimedia*, MUM 2018, pages 245–252, New York, NY, USA, 2018. ACM.

[16] Y. Abdrabou, K. Kassem, J. Salah, R. El-Gendy, M. Morsy, Y. Abdelrahman, and S. Abdennadher. Exploring the usage of eeg and pupil diameter to detect elicited valence. In *International Conference on Intelligent Human Systems Integration*, pages 287–293. Springer, 2018.

[17] W. Abouelsadat. Lineup: List navigation using edge menu. In *Intelligent Human Computer Interaction: 9th International Conference, IHCI 2017, Evry, France, December 11-13, 2017, Proceedings*, volume 10688, page 86. Springer, 2018.

[18] D. W. Aha. Lazy learning. chapter Lazy Learning, pages 7–10. Kluwer Academic Publishers, Norwell, MA, USA, 1997.

[19] F. Alt, M. Mikusz, S. Schneegass, and A. Bulling. Long-term memorability of cued-recall graphical passwords with saliency masks. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, MUM'16, New York, NY, USA, 2016. ACM.

[20] F. Alt, S. Schneegass, A. S. Shirazi, M. Hassib, and A. Bulling. Graphical passwords in the wild: Understanding how users choose pictures and passwords in image-based authentication schemes. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pages 316–322, New York, NY, USA, 2015. ACM.

[21] F. Amon, A. Hamins, N. Bryner, and J. Rowe. Meaningful performance evaluation conditions for fire service thermal imaging cameras. *Fire Safety Journal*, 43(8):541–550, 2008.

[22] F. Amon and C. Pearson. Thermal imaging in firefighting and thermography applications. *Radiometric Temperature Measurements: II. Applications*, 43:279–331, 2009.

[23] J. Anand, T. A. Flora, and A. S. Philip. Finger-vein based biometric security system. *International Journal of Research in Engineering and Technology eISSN*, pages 2319–1163, 2013.

[24] P. Andriotis, T. Tryfonas, G. Oikonomou, and C. Yildiz. A pilot study on the security of pattern screen-lock methods and soft side channel attacks. In *Proceedings of the Sixth ACM Conference on Security and Privacy in*

*Wireless and Mobile Networks*, WiSec '13, pages 1–6, New York, NY, USA, 2013. ACM.

[25] J. R. Angell. *Psychology: An introductory study of the structure and function of human consciousness*. H. Holt, 1910.

[26] J. Anthony, L & Wobbrock. A lightweight multistroke recognizer for user interface prototypes. In *Proc. GI*, 2010.

[27] J. Auda, D. Weber, A. Voit, and S. Schneegass. Understanding user preferences towards rule-based notification deferral. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW584:1–LBW584:6, New York, NY, USA, 2018. ACM.

[28] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies*, WOOT'10, pages 1–7, Berkeley, CA, USA, 2010. USENIX.

[29] J. Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2):276, 1982.

[30] P. Beckmann and A. Spizzichino. The scattering of electromagnetic waves from rough surfaces. *Norwood, MA, Artech House, Inc., 1987, 511 p.*, 1, 1987.

[31] H. Bennett and J. Porteus. Relation between surface roughness and specular reflectance at normal incidence. *Journal of the Optical Society of America (1917-1983)*, 51:123, 1961.

[32] B. Berkovitz, C. Kirsch, J. B. Moxham, G. Alusi, and T. Cheesman. 3d head and neck anatomy with special senses and basic neuroanatomy. [dvd-rom]. United Kingdom: Primal Pictures, 2013.

[33] A. Bianchi, I. Oakley, V. Kostakos, and D. S. Kwon. The phone lock: Audio and haptic shoulder-surfing resistant pin entry methods for mobile devices. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '11, pages 197–200, New York, NY, USA, 2011. ACM.

[34] A. Bianchi, I. Oakley, and D. Kwon. Spinlock: A single-cue haptic and audio pin input technique for authentication. In E. Cooper, V. Kryssanov, H. Ogawa, and S. Brewster, editors, *Haptic and Audio Interaction Design*,

volume 6851 of *Lecture Notes in Computer Science*, pages 81–90. Springer Berlin Heidelberg, 2011.

[35] A. Bianchi, I. Oakley, and D. S. Kwon. Counting clicks and beeps: Exploring numerosity based haptic and audio pin entry. *Interacting with Computers*, 24(5):409 – 422, 2012.

[36] J. Biswas and M. Veloso. Depth camera based indoor mobile robot localization and navigation. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1697–1702, May 2012.

[37] A. Blandford, D. Furniss, and S. Makri. Qualitative hci research: Going behind the scenes. *Synthesis Lectures on Human-Centered Informatics*, 9(1):1–115, 2016.

[38] G. Blasko, S. Feiner, and F. Coriand. Exploring interaction with a simulated wrist-worn projection display. In *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, ISWC '05, pages 2–9, Washington, DC, USA, 2005. IEEE Computer Society.

[39] A. E. Bolock, J. Salah, S. Abdennadher, and Y. Abdelrahman. Character computing: challenges and opportunities. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, pages 555–559. ACM, 2017.

[40] A. Bulling. Pervasive attentive user interfaces. *IEEE Computer*, 49(1):94–98, 2016.

[41] A. Bulling, F. Alt, and A. Schmidt. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 3011–3020, New York, NY, USA, 2012. ACM.

[42] D. Buschek, A. De Luca, and F. Alt. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1393–1402, New York, NY, USA, 2015. ACM.

[43] P. J. Camp, J. M. Hudson, R. B. Keldorph, S. Lewis, and E. D. Mynatt. Supporting communication and collaboration practices in safety-critical situations. In *CHI'00 extended abstracts on Human factors in computing systems*, pages 249–250. ACM, 2000.

[44] X. Cao and R. Balakrishnan. Interacting with dynamically defined information spaces using a handheld projector and a pen. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, UIST '06, pages 225–234, New York, NY, USA, 2006. ACM.

[45] X. Cao, C. Forlines, and R. Balakrishnan. Multi-user interaction using handheld projectors. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, UIST '07, pages 43–52, New York, NY, USA, 2007. ACM.

[46] C. Carreiras, A. Lourenço, H. Aidos, H. P. da Silva, and A. L. N. Fred. *Unsupervised Analysis of Morphological ECG Features for Attention Detection*, pages 437–453. Springer International Publishing, Cham, 2016.

[47] K. W. Cho, F. Lin, C. Song, X. Xu, F. Gu, and W. Xu. Thermal handprint analysis for forensic identification using heat-earth mover's distance. In *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–8. IEEE, 2016.

[48] C. J. Cohen, G. Beach, and G. Foulk. A basic hand gesture control system for pc applications. In *Applied Imagery Pattern Recognition Workshop, AIPR 2001 30th*, pages 74–79. IEEE, 2001.

[49] M. Cooper, B. Mikic, and M. Yovanovich. Thermal contact conductance. *International Journal of heat and mass transfer*, 12(3), 1969.

[50] L. Coventry, A. De Angeli, and G. Johnson. Usability and biometric verification at the atm interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 153–160, New York, NY, USA, 2003. ACM.

[51] L. G. Cowan and K. A. Li. Shadowpuppets: supporting collocated interaction with mobile projector phones using hand shadows. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2707–2716, New York, NY, USA, 2011. ACM.

[52] J. Cross and C. Smith. Thermographic imaging of the subcutaneous vascular network of the back of the hand for biometric identification. In *Security Technology, 1995. Proceedings. Institute of Electrical and Electronics Engineers 29th Annual 1995 International Carnahan Conference on*, pages 20–35, Oct 1995.

[53] K. Crowley, A. Sliney, I. Pitt, and D. Murphy. Evaluating a brain-computer interface to categorise human emotional response. In *ICALT*, pages 276–278, 2010.

[54] G. A. K. Daniel P. Huttenlocher and W. J. Rucklidge. Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850–863, Sep 1993.

[55] T. H. Davenport and J. C. Beck. The attention economy. *Ubiquity*, 2001(May), May 2001.

[56] A. De Luca, A. Hang, E. von Zezschwitz, and H. Hussmann. I feel like i'm taking selfies all day!: Towards understanding biometric authentication on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1411–1414, New York, NY, USA, 2015. ACM.

[57] A. De Luca, M. Harbach, E. von Zezschwitz, M.-E. Maurer, B. E. Slawik, H. Hussmann, and M. Smith. Now you see me, now you don't: Protecting smartphone authentication from shoulder surfers. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2937–2946, New York, NY, USA, 2014. ACM.

[58] A. De Luca, E. von Zezschwitz, N. D. H. Nguyen, M.-E. Maurer, E. Rubegni, M. P. Scipioni, and M. Langheinrich. Back-of-device authentication on smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2389–2398, New York, NY, USA, 2013. ACM.

[59] A. De Luca, E. von Zezschwitz, L. Pichler, and H. Hussmann. Using fake cursors to secure on-screen password entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2399–2402, New York, NY, USA, 2013. ACM.

[60] S. Denef, L. Ramirez, and T. Dyrks. Letting tools talk: interactive technology for firefighting. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 4447–4452. ACM, 2009.

[61] K. Dijkstra and H. van Arem. Computer vision: Thermal imaging and infrared. http://www.optotherm.com/emiss-physics.htm, September 2012. Accessed: 2013-10-11.

[62] T. Dingler, A. Schmidt, and T. Machulla. Building cognition-aware systems: A mobile toolkit for extracting time-of-day fluctuations of cognitive performance. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3):47:1–47:15, Sept. 2017.

[63] A. Draz, S. Abdennadher, and Y. Abdelrahman. Kodr: A customizable learning platform for computer science education. In *European Conference on Technology Enhanced Learning*, pages 579–582. Springer, 2016.

[64] S. Dreiseitl and L. Ohno-Machado. Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, 35(5):352 – 359, 2002.

[65] R. O. Duda and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[66] M. Eiband, M. Khamis, E. von Zezschwitz, H. Hussmann, and F. Alt. Understanding shoulder surfing in the wild: Stories from users and observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, New York, NY, USA, 2017. ACM.

[67] M. Eiband, E. von Zezschwitz, D. Buschek, and H. Hußmann. My scrawl hides it all: Protecting text messages against shoulder surfing with handwritten fonts. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2041–2048, New York, NY, USA, 2016. ACM.

[68] A. Eidels, J. T. Townsend, and D. Algom. Comparing perception of stroop stimuli in focused versus divided attention paradigms: Evidence for dramatic processing differences. *Cognition*, 114(2):129 – 150, 2010.

[69] A. El Bolock, J. Salah, Y. Abdelrahman, C. Herbert, and S. Abdennadher. Character computing: Computer science meets psychology. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, pages 557–562, New York, NY, USA, 2018. ACM.

[70] P. Elagroudy, Y. Abdelrahman, S. Faltaous, S. Schneegass, and H. Davis. Workshop on amplified and memorable food interactions. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, pages 551–555, New York, NY, USA, 2018. ACM.

[71] M. ElKomy, Y. Abdelrahman, M. Funk, T. Dingler, A. Schmidt, and S. Abdennadher. Abbas: An adaptive bio-sensors based assistive system. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 2543–2550, New York, NY, USA, 2017. ACM.

[72] D. C. Engelbart. Augmenting human intellect: a conceptual framework (1962). *Technical Report for the Air Force Office of Scientific Research*, 1962.

[73] V. Engert, A. Merla, J. A. Grant, D. Cardone, A. Tusche, and T. Singer. Exploring the use of thermal infrared imaging in human stress research. *PloS one*, 9(3):e90782, 2014.

[74] C. W. Eriksen and J. E. Hoffman. Temporal and spatial characteristics of selective encoding from visual displays. *Perception & psychophysics*, 12(2):201–204, 1972.

[75] S. H. Fairclough. Fundamentals of physiological computing. *Interact. Comput.*, 21(1-2):133–145, Jan. 2009.

[76] J. Falkinger. Limited attention as a scarce resource in information-rich economies. *The Economic Journal*, 118(532):1596–1620.

[77] J. Falkinger. Attention economies. *Journal of Economic Theory*, 133(1):266–294, 2007.

[78] K. Fan, J. Huber, S. Nanayakkara, and M. Inami. Spidervision: extending the human field of view for augmented awareness. In *Proceedings of the 5th Augmented Human International Conference*, page 49. ACM, 2014.

[79] K. Fan, J.-M. Seigneur, S. Nanayakkara, and M. Inami. Electrosmog visualization through augmented blurry vision. In *Proceedings of the 7th Augmented Human International Conference 2016*, page 35. ACM, 2016.

[80] I. Fernández-Cuevas, J. C. B. Marins, J. A. Lastras, P. M. G. Carmona, S. P. Cano, M. Á. García-Concepción, and M. Sillero-Quintana. Classification of factors influencing the use of infrared thermography in humans: A review. *Infrared Physics & Technology*, 71:28–55, 2015.

[81] R. C. Folweiler and W. J. Mallio. Thermal radiation characteristics of transparent semi-transparent and translucent materials under non-isothermal conditions. Technical report, DTIC Document, 1964.

[82] M. Frutos-Pascual and B. Garcia-Zapirain. Assessing visual attention using eye tracking sensors in intelligent cognitive therapies based on serious games. *Sensors*, 15(5):11092–11117, 2015.

[83] I. Fujitsu. Palmsecure-sl, 2012.

[84] M. Funk, T. Dingler, J. Cooper, and A. Schmidt. Stop helping me - i'm bored!: Why assembly assistance needs to be adaptive. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, pages 1269–1273, New York, NY, USA, 2015. ACM.

[85] M. Funk, T. Dingler, J. Cooper, and A. Schmidt. Stop helping me-i'm bored!: why assembly assistance needs to be adaptive. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 1269–1273. ACM, 2015.

[86] M. Funk, S. Schneegass, M. Behringer, N. Henze, and A. Schmidt. An interactive curtain for media usage in the shower. In *Proceedings of the 4th International Symposium on Pervasive Displays*, PerDis '15, pages 225–231, New York, NY, USA, 2015. ACM.

[87] J. M. Gardiner and A. J. Parkin. Attention and recollective experience in recognition memory. *Memory & Cognition*, 18(6):579–583, Nov 1990.

[88] B. Gaver, T. Dunne, and E. Pacenti. Design: cultural probes. *interactions*, 6(1):21–29, 1999.

[89] M. Germain, M. Jobin, and M. Cabanac. The effect of face fanning during recovery from exercise hyperthermia. *Canadian journal of physiology and pharmacology*, 65(1):87–91, 1987.

[90] N. Goyal. Comet: Collaboration in mobile environments by twisting. In *Supplementary Proceedings of the 11th European Conference on Computer Supported Cooperative Work*, volume 29, 2009.

[91] N. Goyal. Designing for collaborative sensemaking: Using expert & non-expert crowd. *CoRR*, abs/1511.06053, 2015.

[92] N. Goyal and S. R. Fussell. Designing for collaborative sensemaking: Leveraging human cognition for complex tasks. *arXiv preprint arXiv:1511.05737*, 2015.

[93] N. Goyal and S. R. Fussell. Intelligent interruption management using electro dermal activity based physiological sensor on collaborative sensemaking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):16, 2017.

[94] A. Greaves and E. Rukzio. Evaluation of picture browsing using a projector phone. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, MobileHCI '08, pages 351–354, New York, NY, USA, 2008. ACM.

[95] E. Grönvall, J. Fritsch, and A. Vallgårda. Feltradio: Sensing and making sense of wireless traffic. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pages 829–840. ACM, 2016.

[96] J. Gugenheimer, A. De Luca, H. Hess, S. Karg, D. Wolf, and E. Rukzio. Colorsnakes: Using colored decoys to secure authentication in sensitive contexts. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pages 274–283, New York, NY, USA, 2015. ACM.

[97] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer. Knn model-based approach in classification. In R. Meersman, Z. Tari, and D. C. Schmidt, editors, *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 986–996, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[98] M. Haak, S. Bos, S. Panic, and L. Rothkrantz. Detecting stress using eye blinks and brain activity from EEG signals. *Proceeding of the 1st driver car interaction and interface (DCII 2008)*, pages 35–60, 2009.

[99] M. Harbach, A. De Luca, and S. Egelman. The anatomy of smartphone unlocking: A field study of android lock screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4806–4817, New York, NY, USA, 2016. ACM.

[100] C. Harrison, H. Benko, and A. D. Wilson. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 441–450, New York, NY, USA, 2011. ACM.

[101] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. volume 50, pages 904–908, 2006.

[102] M. Hassib, S. Schneegass, P. Eiglsperger, N. Henze, A. Schmidt, and F. Alt. Engagemeter: A system for implicit audience engagement sensing using electroencephalography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 5114–5119, New York, NY, USA, 2017. ACM.

[103] E. Hayashi, S. Das, S. Amini, J. Hong, and I. Oakley. Casa: Context-aware scalable authentication. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, pages 3:1–3:10, New York, NY, USA, 2013. ACM.

[104] J. A. Healey and R. W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 6(2):156–166, 2005.

[105] N. Henze, A. Löcken, S. Boll, T. Hesselmann, and M. Pielot. Free-hand gestures for music playback: deriving gestures with a user-centred process. In *Proceedings of the Conference on Mobile and Ubiquitous Multimedia*, pages 16:1–16:10, 2010.

[106] J. Hernandez, P. Paredes, A. Roseway, and M. Czerwinski. Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 51–60. ACM, 2014.

[107] O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz. Interactions in the air: adding further depth to interactive tabletops. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, pages 139–148, New York, NY, USA, 2009. ACM.

[108] J. E. Hoffman and B. Subramaniam. The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, 57(6):787–795, Jan 1995.

[109] C. Holz and F. R. Bentley. On-demand biometrics: Fast cross-device authentication. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3761–3766, New York, NY, USA, 2016. ACM.

[110] C. Holz, S. Buthpitiya, and M. Knaust. Bodyprint: Biometric user identification on mobile devices using the capacitive touchscreen to scan body parts. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3011–3014, New York, NY, USA, 2015. ACM.

[111] C. Holz and M. Knaust. Biometric touch sensing: Seamlessly augmenting each touch with continuous authentication. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pages 303–312, New York, NY, USA, 2015. ACM.

[112] S. M. Hosni, M. E. Gadallah, S. F. Bahgat, and M. S. AbdelWahab. Classification of eeg signals using different feature extraction techniques for mental-task bci. In *Computer Engineering & Systems, 2007. ICCES'07. International Conference on*, pages 220–226. IEEE, 2007.

[113] J. R. Howell, M. P. Menguc, and R. Siegel. *Thermal radiation heat transfer*. CRC press, 2010.

[114] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderbäck. Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 17–24, New York, NY, USA, 2003. ACM.

[115] C. S. Ikehara and M. E. Crosby. Assessing cognitive load with physiological sensors. In *Proceedings of the 38th annual hawaii international conference on system sciences*, pages 295a–295a. IEEE, 2005.

[116] S. Ioannou, S. Ebisch, T. Aureli, D. Bafunno, H. A. Ioannides, D. Cardone, B. Manini, G. L. Romani, V. Gallese, and A. Merla. The autonomic signature of guilt in children: a thermal infrared imaging study. *PloS one*, 8(11):e79440, 2013.

[117] S. Ioannou, V. Gallese, and A. Merla. Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology*, 51(10):951–963, 2014.

[118] D. Iwai and K. Sato. Heat sensation in image creation with thermal vision. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, ACE '05, pages 213–216, New York, NY, USA, 2005. ACM.

[119] D. Iwai and K. Sato. Limpid desk: see-through access to disorderly desktop in projection-based mixed reality. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 112–115. ACM, 2006.

[120] S. Izadi, S. Hodges, S. Taylor, D. Rosenfeld, N. Villar, A. Butler, and J. Westhues. Going beyond the display: a surface technology with an electronically switchable diffuser. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 269–278. ACM, 2008.

[121] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 559–568, New York, NY, USA, 2011. ACM.

[122] G. Jacucci, A. Spagnolli, J. Freeman, and L. Gamberini. *Symbiotic Interaction: A Critical Definition and Comparison to other Human-Computer Paradigms*, pages 3–20. Springer International Publishing, Cham, 2014.

[123] A. Jain, L. Hong, and S. Pankanti. Biometric identification. *Commun. ACM*, 43(2):90–98, Feb. 2000.

[124] S. Jenkins and R. Brown. A correlational analysis of human cognitive activity using infrared thermography of the supraorbital region, frontal eeg and self-report of core affective state. QIRT, 2014.

[125] X. Jiang, J. I. Hong, L. A. Takayama, and J. A. Landay. Ubiquitous computing for firefighters: Field studies and prototypes of large displays for incident command. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 679–686. ACM, 2004.

[126] D. Kahneman. *Attention and Effort*, volume 1063. Prentice-Hall Englewood Cliffs, NJ, 1973.

[127] S. K. Kane, D. Avrahami, J. O. Wobbrock, B. Harrison, A. D. Rea, M. Philipose, and A. LaMarca. Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pages 129–138. ACM, 2009.

[128] J. Kang and K. Babski-Reeves. Detecting mental workload fluctuation during learning of a novel task using thermography. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(19):1527–1531, 2008.

[129] J. Kang, J. McGinley, G. McFadyen, and K. Babski-Reeves. Determining learning level and effective training times using thermography. In *Proceedings of Army Science Conference, Orlando, Florida, USA*, 2006.

[130] K. Kassem, J. Salah, Y. Abdrabou, M. Morsy, R. ElGendy, Y. Abdelrahman, and S. Abdennadher. Diva: exploring the usage of pupil diameter to elicit valence and arousal. In *Proceedings of the 16th International Conference on MUM. ACM*, 2017.

[131] H. Kataoka, H. Kano, H. Yoshida, A. Saijo, M. Yasuda, and M. Osumi. Development of a skin temperature measuring system for non-contact stress evaluation. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 2, pages 940–943 vol.2, 1998.

[132] K. A. Kerns, K. Eso, and J. Thomson. Investigation of a direct intervention for improving attention in young children with adhd. *Developmental Neuropsychology*, 16(2):273–295, 1999.

[133] M. Khamis, F. Alt, M. Hassib, E. von Zezschwitz, R. Hasholzner, and A. Bulling. Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2156–2164, New York, NY, USA, 2016. ACM.

[134] M. M. Khan, M. Ingleby, and R. D. Ward. Automated facial expression classification and affect interpretation using infrared measurement of facial skin temperature variations. *ACM Trans. Auton. Adapt. Syst.*, 1(1):91–113, Sept. 2006.

[135] D. Kimber, J. Vaughan, and E. Rieffel. Augmented perception through mirror worlds. In *Proceedings of the 2nd Augmented Human International Conference*, page 36. ACM, 2011.

[136] C. Kittel and H. Kroemer. Thermal physics 2nd ed, 1980.

[137] H. Koike, Y. Sato, and Y. Kobayashi. Integrating paper and digital information on enhanceddesk: a method for realtime finger tracking on an augmented desk system. *ACM Trans, on Computer-Human Interaction*, 8(4), 2001.

[138] S. G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B. R. Abidi, A. Koschan, M. Yi, and M. A. Abidi. Multiscale fusion of visible and thermal ir

images for illumination-invariant face recognition. *International Journal of Computer Vision*, 71(2):215–233, 2007.

[139] T. Kosch, Y. Abdelrahman, M. Funk, and A. Schmidt. One size does not fit all: Challenges of providing interactive worker assistance in industrial settings. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 1006–1011, New York, NY, USA, 2017. ACM.

[140] N. Kosmyna, C. Morris, U. Sarawgi, and P. Maes. Attentivu: A biofeedback system for real-time monitoring and improvement of engagement. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pages VS07:1–VS07:2, New York, NY, USA, 2019. ACM.

[141] A. Kumar and K. V. Prathyusha. Personal authentication using hand vein triangulation and knuckle shape. *Image Processing, IEEE Transactions on*, 18(9):2127–2136, Sept 2009.

[142] K. Kuraoka and K. Nakamura. The use of nasal skin temperature measurements in studying emotion in macaque monkeys. *Physiology & behavior*, 102(3):347–355, 2011.

[143] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy. Stress detection from speech and galvanic skin response signals. In *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*, pages 209–214. IEEE, 2013.

[144] P. Kuyper. The cocktail party effect. *Audiology*, 11(5-6):277–282, 1972.

[145] M. Kyng, E. T. Nielsen, and M. Kristensen. Challenges in designing interactive systems for emergency response. In *Proceedings of the 6th conference on Designing Interactive systems*, pages 301–310. ACM, 2006.

[146] E. Labbé, N. Schmidt, J. Babin, and M. Pharr. Coping with stress: the effectiveness of different types of music. *Applied psychophysiology and biofeedback*, 32(3-4):163–168, 2007.

[147] M. J. M. Lamers, A. Roelofs, and I. M. Rabeling-Keus. Selective attention and response set in the stroop task. *Memory & Cognition*, 38(7):893–904, Oct 2010.

[148] J. Landgren. Making action visible in time-critical work. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 201–210. ACM, 2006.

[149] J. Landgren and U. Nulden. A study of emergency response work: patterns of mobile phone interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1323–1332. ACM, 2007.

[150] E. Larson, G. Cohn, S. Gupta, X. Ren, B. Harrison, D. Fox, and S. Patel. Heatwave: thermal imaging for surface user interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2565–2574, New York, NY, USA, 2011. ACM.

[151] M. H. A. Latif, H. M. Yusof, S. N. Sidek, and N. Rusli. Implementation of glcm features in thermal imaging for human affective state detection. *Procedia Computer Science*, 76:308 – 315, 2015. 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IEEE IRIS2015).

[152] S. Leroy. Why is it so hard to do my work? the challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes*, 109(2):168 – 181, 2009.

[153] R. W. Levenson. Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. 1988.

[154] Y. Li, X. Li, M. Ratcliffe, L. Liu, Y. Qi, and Q. Liu. A real-time eeg-based bci system for attention recognition in ubiquitous environment. In *Proceedings of 2011 International Workshop on Ubiquitous Affective Awareness and Intelligent Interaction*, UAAII '11, pages 33–40, New York, NY, USA, 2011. ACM.

[155] J. C. Licklider. Man-computer symbiosis. *IRE transactions on human factors in electronics*, (1):4–11, 1960.

[156] L. Lischke, S. Mayer, A. Preikschat, M. Schweizer, B. Vu, P. W. Woźniak, and N. Henze. Understanding large display environments: Contextual inquiry in a control room. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW134:1–LBW134:6, New York, NY, USA, 2018. ACM.

[157] D. Liu, B. Dong, X. Gao, and H. Wang. Exploiting eye tracking for smartphone authentication. In *Proc. of ACNS '15*.

[158] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu. Recognizing the degree of human attention using eeg signals from mobile sensors. *Sensors*, 13(8):10273–10286, 2013.

[159] J. T. Lussier and S. Thrun. Automatic calibration of rgbd and thermal cameras. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 451–458, Sept 2014.

[160] J. Ma, D. Lei, X. Jin, X. Du, F. Jiang, F. Li, Y. Zhang, and X. Shen. Compensatory brain activation in children with attention deficit/hyperactivity disorder during a simplified go/no-go task. *Journal of neural transmission*, 119(5):613–619, 2012.

[161] W. E. MacKay. Is paper safer? the role of paper flight strips in air traffic control. *ACM Trans. Comput.-Hum. Interact.*, 6(4):311–340, Dec. 1999.

[162] L. Maestri and R. Wakkary. Understanding repair as a creative process of everyday design. In *Proceedings of the 8th ACM conference on Creativity and cognition*, pages 81–90. ACM, 2011.

[163] E. Makled, Y. Abdelrahman, N. Mokhtar, V. Schwind, S. Abdennadher, and A. Schmidt. I like to move it: Investigating the effect of head and body movement of avatars in vr on user's perception. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW099:1–LBW099:6, New York, NY, USA, 2018. ACM.

[164] M. Mancas, V. P. Ferrera, N. Riche, and J. G. Taylor. *From Human Attention to Computational Attention: A Multidisciplinary Approach*, volume 10. Springer, 2016.

[165] B. Manini, D. Cardone, S. J. Ebisch, D. Bafunno, T. Aureli, and A. Merla. Mom feels what her child feels: thermal signatures of vicarious autonomic response while watching children in a stressful situation. *Front. Hum. Neurosci*, 7, 2013.

[166] C. Manresa, J. Varona, R. Mas, and F. Perales. Hand tracking and gesture recognition for human-computer interaction. *Electronic letters on computer vision and image analysis*, 5(3):96–104, 2005.

[167] G. Mark, M. Czerwinski, and S. T. Iqbal. Effects of individual differences in blocking workplace distractions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 92. ACM, 2018.

[168] K. Matsumoto, W. Nakagawa, H. Saito, M. Sugimoto, T. Shibata, and S. Yachida. Ar visualization of thermal 3d model by hand-held cameras. In *SciTePress*, 2015.

[169] T. McCaffrey, R. McCook, and R. Wurster. Effect of head skin temperature on tympanic and oral temperature in man. *Journal of Applied Physiology*, 39(1):114–118, 1975.

[170] A. Merla and G. L. Romani. Thermal signatures of emotional arousal: a functional infrared imaging study. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 247–249. IEEE, 2007.

[171] A. F. Mirsky, B. J. Anthony, C. C. Duncan, M. B. Ahearn, and S. G. Kellam. Analysis of the elements of attention: A neuropsychological approach. *Neuropsychology Review*, 2(2):109–145, Jun 1991.

[172] P. Mistry, P. Maes, and L. Chang. Wuw - wear ur world: a wearable gestural interface. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 4111–4116, New York, NY, USA, 2009. ACM.

[173] M. Moisala, V. Salmela, E. Salo, S. Carlson, V. Vuontela, O. Salonen, and K. Alho. Brain activity during divided and selective attention to auditory and visual sentence comprehension tasks. *Frontiers in Human Neuroscience*, 9:86, 2015.

[174] K. Mowery, S. Meiklejohn, and S. Savage. Heat of the moment: Characterizing the efficacy of thermal camera-based attacks. In *Proceedings of the 5th USENIX Conference on Offensive Technologies*, WOOT'11, Berkeley, CA, USA, 2011. USENIX Association.

[175] S. Murugappan, Vinayak, N. Elmqvist, and K. Ramani. Extended multitouch: recovering touch posture and differentiating users using a depth camera. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 487–496, New York, NY, USA, 2012. ACM.

[176] A. M. Naghsh and C. R. Roast. Designing user interaction with robots swarms in emergency settings. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, pages 519–522. ACM, 2008.

[177] M. Naveh-Benjamin, J. Guez, Y. Hara, M. S. Brubaker, and I. Lowenschuss-Erlich. The effects of divided attention on encoding processes under incidental and intentional learning instructions: Underlying mechanisms? *Quarterly Journal of Experimental Psychology*, 67(9):1682–1696, 2014. PMID: 24283628.

[178] J. Newn, F. Allison, E. Velloso, and F. Vetere. Looks can be deceiving: Using gaze visualisation to predict and mislead opponents in strategic gameplay. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 261:1–261:12, New York, NY, USA, 2018. ACM.

[179] J. Newn, E. Velloso, F. Allison, Y. Abdelrahman, and F. Vetere. Evaluating real-time gaze representations to infer intentions in competitive turn-based strategy games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '17, pages 541–552, New York, NY, USA, 2017. ACM.

[180] E. Niforatos, I. Elhart, A. Fedosov, and M. Langheinrich. s-helmet: A ski helmet for augmenting peripheral perception. In *Proceedings of the 7th Augmented Human International Conference 2016*, page 45. ACM, 2016.

[181] E. Niforatos, A. Fedosov, I. Elhart, and M. Langheinrich. Augmenting skiers' peripheral perception. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 114–121. ACM, 2017.

[182] A. Ninio and D. Kahneman. Reaction time in focused and in divided attention. *Journal of Experimental Psychology*, 103(3):394, 1974.

[183] I. Oakley and A. Bianchi. Multi-touch passwords for mobile device access. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 611–612, New York, NY, USA, 2012. ACM.

[184] I. Oakley and A. Bianchi. Keeping Secrets from Friends. *Archives of Design Research*, 27:49–62, 2014.

[185] T. Oka, K. Oka, and T. Hori. Mechanisms and mediators of psychological stress-induced rise in core temperature. *Psychosomatic medicine*, 63(3):476–486, 2001.

[186] C. K. Or and V. G. Duffy. Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. *Occupational Ergonomics*, 7(2):83–94, 2007.

[187] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[188] A. Parnandi, B. Ahmed, E. Shipp, and R. Gutierrez-Osuna. Chill-out: Relaxation training through respiratory biofeedback in a mobile casual game. In *Mobile Computing, Applications, and Services*, pages 252–260. Springer, 2013.

[189] A. Parnandi, Y. Son, and R. Gutierrez-Osuna. A control-theoretic approach to adaptive physiological games. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 7–12. IEEE, 2013.

[190] K. Parsons. Contact between human skin & hot surfaces equivalent contact temperature. In *Proc. ICEE*, 1992.

[191] I. Pavlidis and J. Levine. Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine*, 21(6):56–64, 2002.

[192] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddharaju, R. Joseph, A. Mandapati, B. Dunkin, et al. Fast by nature-how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports*, 2, 2012.

[193] M. K. Petersen, C. Stahlhut, A. Stopczynski, J. E. Larsen, and L. K. Hansen. Smartphones get emotional: mind reading images and reconstructing the neural sources. In *International Conference on Affective Computing and Intelligent Interaction*, pages 578–587. Springer, 2011.

[194] B. Pfleging, D. K. Fekety, A. Schmidt, and A. L. Kun. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '16, New York, NY, USA, May 2016. ACM.

[195] Q.-C. Pham, L. Gond, J. Begard, N. Allezard, and P. Sayd. Real-time posture analysis in a crowd using thermal imaging. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[196] R. W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, USA, 1997.

[197] M. Planck. Über eine verbesserung der wienschen spectralgleichung. *Verhandl. Dtsc. Phys. Ges.*, 2:202, 1900.

[198] A. P. Pons and P. Polak. Understanding user perspectives on biometric technology. *Commun. ACM*, 51(9):115–118, Sept. 2008.

[199] M. I. Posner. Orienting of attention. *Quarterly journal of experimental psychology*, 32(1):3–25, 1980.

[200] M. I. Posner and S. E. Petersen. The attention system of the human brain. *Annual Review of Neuroscience*, 13(1):25–42, 1990. PMID: 2183676.

[201] R. Prasanna, L. Yang, and M. King. Guidance for developing human–computer interfaces for supporting fire emergency response. *Risk Management*, 15(3):155–179, 2013.

[202] R. Prasanna, L. Yang, and M. King. Human–computer interaction for supporting fire emergency first responders. In *Proceedings of the people and the planet 2013 conference: Transforming the future, Melbourne.*, 2013.

[203] C. Puri, L. Olson, I. Pavlidis, J. Levine, and J. Starren. Stresscam: non-contact measurement of users' emotional states through thermal imaging. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1725–1728. ACM, 2005.

[204] P. J. Qualls and P. W. Sheehan. Role of the feedback signal in electromyograph biofeedback: The relevance of attention. *Journal of Experimental Psychology: General*, 110(2):204, 1981.

[205] T. J. Quinn and J. Martin. A radiometric determination of the stefan-boltzmann constant and thermodynamic temperatures between-40 c and+ 100 c. *Phil. Trans. R. Soc. Lond. A*, 316(1536):85–189, 1985.

[206] B. A. Rajoub and R. Zwiggelaar. Thermal facial analysis for deception detection. *IEEE Transactions on Information Forensics and Security*, 9(6):1015–1023, June 2014.

[207] R. Ray. The theory and practice of safe handling temperatures. *Applied ergonomics*, 15(1), 1984.

[208] E. F. J. Ring and K. Ammer. Infrared thermal imaging in medicine. *Physiological Measurement*, 33(3):R33, 2012.

[209] E. Rukzio, P. Holleis, and H. Gellersen. Personal projectors for pervasive computing. *IEEE Pervasive Computing*, 11(2):30–37, Apr. 2012.

[210] A. Sahami Shirazi, Y. Abdelrahman, N. Henze, S. Schneegass, M. Khalilbeigi, and A. Schmidt. Exploiting thermal reflection for interactive systems. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 3483–3492, New York, NY, USA, 2014. ACM.

[211] J. Salah, Y. Abdelrahman, Y. Abdrabou, K. Kassem, and S. Abdennadher. Exploring the usage of commercial bio-sensors for multitasking detection. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, pages 265–277, New York, NY, USA, 2018. ACM.

[212] J. Salah, Y. Abdelrahman, A. Dakrouni, and S. Abdennadher. Judged by the cover: Investigating the effect of adaptive game interface on the learning experience. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, pages 215–225, New York, NY, USA, 2018. ACM.

[213] J. Salah, S. Abdennadher, C. Sabty, and Y. Abdelrahman. Super alpha: Arabic alphabet learning serious game for children with learning disabilities. In *Joint International Conference on Serious Games*, pages 104–115. Springer, 2016.

[214] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, pages 71–78, New York, NY, USA, 2000. ACM.

[215] A. Schmidt. Augmenting human intellect and amplifying perception and cognition. *IEEE Pervasive Computing*, 16(1):6–10, 2017.

[216] A. Schmidt. Understanding and researching through making: A plea for functional prototypes. *interactions*, 24(3):78–81, Apr. 2017.

[217] A. Schmidt, L. Terrenghi, and P. Holleis. Methods and guidelines for the design and development of domestic ubiquitous computing applications. *Pervasive and Mobile Computing*, 3(6):721–738, 2007.

[218] D. Schmidt, D. Molyneaux, and X. Cao. Picontrol: using a handheld projector for direct control of physical devices through visible light. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 379–388, New York, NY, USA, 2012. ACM.

[219] S. Schneegass, B. Pfleging, N. Broy, F. Heinrich, and A. Schmidt. A data set of real world driving to assess driver workload. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 150–157. ACM, 2013.

[220] S. Schneegass, F. Steimle, A. Bulling, F. Alt, and A. Schmidt. Smudgesafe: Geometric image transformations for smudge-resistant user authentication. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pages 775–786, New York, NY, USA, 2014. ACM.

[221] J. Selvaraj, M. M, R. Nagarajan, and K. Wan. Physiological signals based human emotion recognition: a review. 03 2011.

[222] N. Sharma and T. Gedeon. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer methods and programs in biomedicine*, 108(3):1287–1301, 2012.

[223] D. Shastri, A. Merla, P. Tsiamyrtzis, and I. Pavlidis. Imaging facial signs of neurophysiological responses. *IEEE Transactions on Biomedical Engineering*, 56(2):477–484, 2009.

[224] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis. Perinasal imaging of physiological stress and its affective potential. *IEEE Transactions on Affective Computing*, 3(3):366–378, 2012.

[225] R. Shewbridge, A. Hurst, and S. K. Kane. Everyday making: identifying future uses for 3d printing in the home. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 815–824. ACM, 2014.

[226] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen. Galvanic skin response (gsr) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 2651–2656. ACM, 2007.

[227] A. S. Shirazi, M. Funk, F. Pfleiderer, H. Glück, and A. Schmidt. Mediabrain: Annotating videos based on brain-computer interaction. In *Mensch & Computer*, pages 263–272, 2012.

[228] A. S. Shirazi, M. Hassib, N. Henze, A. Schmidt, and K. Kunze. What's on your mind?: mental task awareness using single electrode brain computer interfaces. In *Proceedings of the 5th Augmented Human International Conference*, page 45. ACM, 2014.

[229] R. Sinha, W. R. Lovallo, and O. A. Parsons. Cardiovascular differentiation of emotions. *Psychosomatic Medicine*, 54(4):422–435, 1992.

[230] M. M. Sohlberg and C. A. Mateer. Effectiveness of an attention-training program. *Journal of Clinical and Experimental Neuropsychology*, 9(2):117–130, 1987. PMID: 3558744.

[231] S. Sonkamble, D. R. Thool, and B. Sonkamble. Survey of biometric recognition systems and their applications. *Journal of Theoretical & Applied Information Technology*, 11, 2010.

[232] E. M. Sparrow. Radiation heat transfer. 1961.

[233] N. Srivastava, J. Newn, and E. Velloso. Combining low and mid-level gaze features for desktop activity recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4):189:1–189:27, Dec. 2018.

[234] J. Steil, P. Müller, Y. Sugano, and A. Bulling. Forecasting user attention during everyday mobile interactions using device-integrated and wearable sensors. *arXiv preprint arXiv:1801.06011*, 2018.

[235] J. Stemberger, R. S. Allison, and T. Schnell. Thermal imaging as a way to classify cognitive workload. In *Computer and Robot Vision (CRV), 2010 Canadian Conference on*, pages 231–238. IEEE, 2010.

[236] J. R. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General*, 121(1):15, 1935.

[237] W. Sturm, K. Willmes, B. Orgass, and W. Hartje. Do specific attention deficits need specific training? *Neuropsychological Rehabilitation*, 7(2):81–103, 1997.

[238] E. A. Styles. *The Psychology of Attention*. Psychology Press, 1997.

[239] J. Sweller, P. Ayres, and S. Kalyuga. *Measuring Cognitive Load*, pages 71–85. Springer New York, New York, NY, 2011.

[240] B. Tag, R. Mannschreck, K. Sugiura, G. Chernyshov, N. Ohta, and K. Kunze. Facial thermography for attention tracking on smart eyewear: An initial study. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA ’17, pages 2959–2966, New York, NY, USA, 2017. ACM.

[241] C. D. Technologies. Glass material information. http://www.sydor.com/wp-content/uploads/Corning-EAGLE-XG-Display-Glass.pdf, 2006. Accessed September 19, 2016.

[242] S. Uellenbeck, M. Dürmuth, C. Wolf, and T. Holz. Quantifying the security of graphical passwords: The case of android unlock patterns. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, pages 161–172, New York, NY, USA, 2013. ACM.

[243] J. Van Baar, P. Beardsley, M. Pollefeys, and M. Gross. Sensor fusion for depth estimation, including tof and thermal sensors. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pages 472–478. IEEE, 2012.

[244] G. van Daalen, T. M. Willemsen, K. Sanders, and M. J. van Veldhoven. Emotional exhaustion and mental health problems among employees doing âpeople workâ: The impact of job demands, job resources and family-to-work conflict. *International archives of occupational and environmental health*, 82(3):291–303, 2009.

[245] B. M. Velichkovsky and J. P. Hansen. New technological windows into mind: There is more in eyes and brains for human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 496–503. ACM, 1996.

[246] M. Vidal, J. Turner, A. Bulling, and H. Gellersen. Wearable eye tracking for mental health monitoring. *Computer Communications*, 35(11):1306 – 1311, 2012.

[247] S. Vidas, P. Moghadam, and M. Bosse. 3d thermal mapping of building interiors using an rgb-d and thermal camera. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 2311–2318. IEEE, 2013.

[248] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[249] A. Voit, B. Poppinga, D. Weber, M. Böhmer, N. Henze, S. Gehring, T. Okoshi, and V. Pejovic. Ubittention: Smart & ambient notification and attention management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, pages 1520–1523, New York, NY, USA, 2016. ACM.

[250] M. Vollmer, S. Henke, D. Karstädt, K. Möllmann, and F. Pinno. Identification and suppression of thermal reflections in infrared thermal imaging. In *Proceedings of Inframation*, volume 5, pages 287–298, 2004.

[251] E. von Zezschwitz, A. De Luca, B. Brunkow, and H. Hussmann. Swipin: Fast and secure pin-entry on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1403–1406, New York, NY, USA, 2015. ACM.

[252] E. von Zezschwitz, A. De Luca, and H. Hussmann. *Survival of the Shortest: A Retrospective Analysis of Influencing Factors on Password Composition*, pages 460–467. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[253] E. von Zezschwitz, A. De Luca, P. Janssen, and H. Hussmann. Easy to draw, but hard to trace?: On the observability of grid-based (un)lock patterns. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2339–2342, New York, NY, USA, 2015. ACM.

[254] E. von Zezschwitz, P. Dunphy, and A. De Luca. Patterns in the wild: A field study of the usability of pattern and pin-based authentication on mobile devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, pages 261–270, New York, NY, USA, 2013. ACM.

[255] E. von Zezschwitz, S. Ebbinghaus, H. Hussmann, and A. De Luca. You can't watch this!: Privacy-respectful photo browsing on smartphones. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4320–4324, New York, NY, USA, 2016. ACM.

[256] E. von Zezschwitz, A. Koslow, A. De Luca, and H. Hussmann. Making graphic-based authentication secure against smudge attacks. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI '13, pages 277–286, New York, NY, USA, 2013. ACM.

[257] R. Wakkary and L. Maestri. The resourcefulness of everyday design. In *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*, pages 163–172. ACM, 2007.

[258] J. Wallace, J. McCarthy, P. C. Wright, and P. Olivier. Making design probes work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3441–3450. ACM, 2013.

[259] J.-G. Wang, W.-Y. Yau, A. Suwandy, and E. Sung. Person recognition by fusing palmprint and palm vein images based on laplacianpalm representation. *Pattern Recognition*, 41(5):1514 – 1527, 2008.

[260] L. Wang and G. Leedham. A thermal hand vein pattern verification system. In *Pattern Recognition and Image Analysis*, pages 58–65. Springer, 2005.

[261] Z. Wang, J. Yan, and H. Aghajan. A framework of personal assistant for computer users by analyzing video stream. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*, page 14. ACM, 2012.

[262] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063, 1988.

[263] D. Weber, A. S. Shirazi, S. Gehring, N. Henze, B. Poppinga, M. Pielot, and T. Okoshi. Smarttention, please!: 2nd workshop on intelligent attention management on mobile devices. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '16, pages 914–917, New York, NY, USA, 2016. ACM.

[264] M. Weiser. The computer for the 21st century. *Scientific american*, 265(3):94–104, 1991.

[265] O. Wiese and V. Roth. See you next time: A model for modern shoulder surfers. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '16, pages 453–464, New York, NY, USA, 2016. ACM.

[266] K. D. Willis, I. Poupyrev, and T. Shiratori. Motionbeam: a metaphor for character interaction with handheld projectors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1031–1040, New York, NY, USA, 2011. ACM.

[267] A. D. Wilson. Touchlight: an imaging touch screen and display for gesture-based interaction. In *Proceedings of the 6th international conference on Multimodal interfaces*, ICMI '04, pages 69–76, New York, NY, USA, 2004. ACM.

[268] A. D. Wilson. Playanywhere: a compact interactive tabletop projection-vision system. In *Proceedings of the 18th annual ACM*

*symposium on User interface software and technology*, pages 83–92. ACM, 2005.

[269] A. D. Wilson. Depth sensing video cameras for 3d tangible tabletop interaction. In *In Proceedings of Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, 2007.

[270] A. D. Wilson. Using a depth camera as a touch sensor. In *Proceedings of the Conference on Interactive Tabletops and Surfaces*, pages 69–72, 2010.

[271] A. D. Wilson and H. Benko. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pages 273–282, New York, NY, USA, 2010. ACM.

[272] C. Winkler, K. Pfeuffer, and E. Rukzio. Investigating mid-air pointing interaction for projector phones. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, ITS '12, pages 85–94, New York, NY, USA, 2012. ACM.

[273] C. Winkler, C. Reinartz, D. Nowacka, and E. Rukzio. Interactive phone call: synchronous remote collaboration and projected interactive surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 61–70, New York, NY, USA, 2011. ACM.

[274] J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: a recognizer for user interface prototypes. In *Proc. UIST*, 2007.

[275] K. Wolf, E. Abdelhady, Y. Abdelrahman, T. Kubitza, and A. Schmidt. mesch: tools for interactive exhibitions. In *Proceedings of the Conference on Electronic Visualisation and the Arts*, pages 261–269. BCS Learning & Development Ltd., 2015.

[276] K. Wolf, Y. Abdelrahman, T. Kubitza, and A. Schmidt. Proxemic zones of exhibits and their manipulation using floor projection. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays*, pages 33–37. ACM, 2016.

[277] K. Wolf, Y. Abdelrahman, M. Landwehr, G. Ward, and A. Schmidt. How to browse through my large video data: face recognition & prioritizing for lifelog video. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, pages 169–173. ACM, 2016.

[278] K. Wolf, Y. Abdelrahman, D. Schmid, T. Dingler, and A. Schmidt. Effects of camera position and media type on lifelogging images. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pages 234–244. ACM, 2015.

[279] K. Wolf, S. Schneegass, N. Henze, D. Weber, V. Schwind, P. Knierim, S. Mayer, T. Dingler, Y. Abdelrahman, T. Kubitza, et al. Tuis in the large: Using paper tangibles with mobile devices. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1579–1584. ACM, 2015.

[280] W. K. Wong, H. L. Lim, C. K. Loo, and W. S. Lim. Home alone faint detection surveillance system using thermal camera. In *Computer Research and Development, 2010 Second International Conference on*, pages 747–751. IEEE, 2010.

[281] W. K. Wong, P. N. Tan, C. K. Loo, and W. S. Lim. An effective surveillance system using thermal camera. In *Signal Acquisition and Processing, 2009. ICSAP 2009. International Conference on*, pages 13–17. IEEE, 2009.

[282] H. M. Wood. *The use of passwords for controlled access to computer resources*, volume 500. US Department of Commerce, National Bureau of Standards, 1977.

[283] R. Wood, C. Han, and P. Kruse. Integrated uncooled infrared detector imaging arrays. In *Solid-State Sensor and Actuator Workshop, 1992. 5th Technical Digest., IEEE*, pages 132–135. IEEE, 1992.

[284] P. W. Wozniak, L. Lischke, S. Mayer, A. Preikschat, M. Schweizer, B. Vu, C. von Molo, and N. Henze. Understanding work in public transport management control rooms. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW ’17 Companion, pages 339–342, New York, NY, USA, 2017. ACM.

[285] R. D. Wright and L. M. Ward. *Orienting of attention*. Oxford University Press, 2008.

[286] L. Yang, Y. Ren, and W. Zhang. 3d depth image analysis for indoor fall detection of elderly people. *Digital Communications and Networks*, 2(1):24–34, 2016.

[287] J. Zagermann, U. Pfeil, and H. Reiterer. Studying eye movements as a basis for measuring cognitive load. In *Extended Abstracts of the 2018 CHI*

*Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW095:1–LBW095:6, New York, NY, USA, 2018. ACM.

[288] R. B. Zajonc, S. T. Murphy, and M. Inglehart. Feeling and facial efference: implications of the vascular theory of emotion. *Psychological review*, 96(3):395, 1989.

[289] Q. Zhou, J. Newn, N. Srivastava, T. Dingler, J. Goncalves, and E. Velloso. Cognitive aid: Task assistance based on mental workload estimation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pages LBW2315:1–LBW2315:6, New York, NY, USA, 2019. ACM.

[290] Z. Zhu, P. Tsiamyrtzis, and I. Pavlidis. Forehead thermal signature extraction in lie detection. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 243–246. IEEE, 2007.

## Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12. Juli 2011, § 8, Abs. 2 Pkt. 5.)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

Stuttgart, den 04.10.2018

Yomna Ali Gamaleldin Abdelrahman