

# **Understanding Human Images in Thermal Infrared Spectrum**

**Brahmastro Kresnaraman**



# *Abstract*

In this day and age, security surveillance systems can be seen almost everywhere. They are expected to work in various environments. Most common cameras work in the visible spectrum, in which the human visual system also operates. However, these cameras face a big problem in less-than-ideal illumination conditions. In the nighttime or in dark areas, thermal cameras are better options.

In regards to the surveillance, the most common target of observation is humans. However, it is difficult for humans to observe and interpret thermal infrared images, and cannot easily identify a person captured in them. This brings us to the main theme of this thesis, namely “Understanding thermal infrared images.” The focal point of this thesis is understanding “Human images” in the thermal infrared spectrum. The task of understanding can be given to a system, and then the system can be used to help people understand thermal infrared images in various applications.

Two research questions that may be asked from human images in the thermal infrared spectrum, “What does the person look like in the visible spectrum?” and “What is the person wearing?” are the foci of the thesis. In order to answer these questions, two approaches to obtain understanding were chosen; visually and semantically. Each understanding is investigated separately on its respective research topic.

The first research topic is the visual understanding, in which face images are reconstructed from the thermal infrared to the visible spectra. This reconstruction is very beneficial in the case of using a thermal infrared camera for surveillance, because humans struggle to recognize faces taken in this spectrum. By reconstructing the face to the visible spectrum, humans can then perform the identification easily. Since the biometric information of a face is unique to a person regardless of the spectrum it is in, the system can infer the correlation of a face between different spectra. Therefore, by analyzing the correlation between these face images taken in both the thermal infrared and the visible spectra, the reconstruction can then be performed based on this relationship.

Experimental results showed the high performance of the proposed method in all evaluations for reconstructing known persons' thermal face image, outperforming all other methods. Additionally, a discussion is also provided on reconstructing unknown person's thermal face image. The proposed method as well as the comparative methods struggled to perform this kind of reconstruction because there was no information of the face in the visible spectrum. This proved the difficulty of the task and warrants further research in the field.

The second research topic is the semantic understanding, specifically on the existence of wearable attributes such as glasses, masks, and hat. The existence of wearable attributes on the face of a person may occlude the identification process. In areas where the identity of a person is important, this unknown person could be a security risk. A preventative measure is useful in this case, hence identifying what attributes people are wearing in advance is important. In order to achieve this, a decomposition approach for recognition is proposed. The approach extracts only the attribute information by taking advantage of the properties of the wearable attributes in the thermal infrared spectrum and uses them for the attribute recognition process. Two versions of the proposed methods are introduced, where the first one was improved upon with a prior knowledge of the region where the attributes are likely to be present.

Experimental results showed the high performance of the proposed method that used the prior knowledge, outperforming all other methods. By visual observation, the resulting images also showed relatively better extracted attributes.

In summary, two systems that can provide insights to face and human upper body images in the thermal infrared spectrum were implemented. Face image reconstruction provides visual understanding of what a face image taken in the thermal infrared spectrum looks like in the visible spectrum. Wearable attribute recognition provides semantic understanding to human upper body images, providing information of various objects worn by the person in an image. There are still many areas to improve upon, but these systems can be useful for security surveillance systems.

# *Acknowledgments*

This dissertation is formally submitted for fulfilling partial requirements for the degree of Doctor of Information Science from Nagoya University. It marks the end of my journey after living in Japan from April 2011 and also serves as the pinnacle of what I have achieved in academic field to date. I have experienced many things during my stay and met many people who have left ever lasting impressions on me.

My sincerest gratitude goes to Prof. Dr. Hiroshi Murase, for accepting me to his laboratory. Only with his continuous encouragement and support throughout my stay that it is possible for me to finish my degree. From being a research student until now, his guidance has made it all possible. Even if I could turn back time, I would not ask for another supervisor.

I would also like to give my gratitude to Prof. Dr. Ichiro Ide for his help on most of my publications, including this dissertation. Despite his busy schedule, he is always lightning fast and meticulous on his checks. With his help, many of my writings have been shaped for the better.

I would like to give a big thank you to Prof. Dr. Daisuke Deguchi and Dr. Yasutomo Kawanishi. Prof. Dr. Daisuke Deguchi helped a lot with his insight on various discussions, whether it is regarding my research or my publications. Dr. Yasutomo Kawanishi has guided me a lot, especially in shaping the road for my second research topic.

I thank Prof. Dr. Kenji Mase for the feedback and comments during the reviewing process.

I also would like to thank my supervisor during my master student years, Prof. Dr. Yoshito Mekada. Without his help and support, I could not graduate and continue my pursuit of higher education in Nagoya University. My gratitude also goes to Prof. Dr. Tomokazu Takahashi who has given me a lot of guidance in making my research path clear.

I would give my special thanks to the secretaries of Murase Laboratory, Mrs. Fumiyo Kaba and Mrs. Hiromi Tanaka for making my life easier with their help on various documents. Also to the other students that I cannot name one by one, thank you for your friendship.

I thank the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) that has supported me in the form of Monbukagakusho scholarship throughout my study as Research Student (2011–2012), Master Student (2012–2014), and PhD. Candidate (2014–2017).

I also thank my family in Indonesia. My parents for they are the reason I was born to this world, for raising me and taking care of me until I could stand by myself. My siblings for their liveliness during the years.

My heartfelt thanks goes to my wife, Laksmi Rahadiani, for giving me strength, love, motivation and happiness. Without her, my life would not be the same.

Finally, I thank Allah S.W.T. for all the blessings and grace that I have received from Him.

# Contents

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgments</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xv</b>
<b>Abbreviations</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Understanding Thermal Infrared Images . . . . .	5
1.2.1 Research Question 1: What does the person look like in the visible spectrum? . . . . .	8
1.2.2 Research Question 2: What is the person wearing? . . . . .	8
1.3 Research Topics . . . . .	10
1.3.1 Topic 1: Visual Understanding from Thermal Infrared Image	11
1.3.2 Topic 2: Semantic Understanding from Thermal Infrared Image . . . . .	13
1.4 Structure of the Thesis . . . . .	16
<b>2 Thermal Infrared Spectrum</b>	<b>19</b>
2.1 Electromagnetic Spectrum . . . . .	19
2.2 Infrared Spectrum . . . . .	21
<b>3 Literature Review</b>	<b>27</b>
3.1 Human Image Processing in the Thermal Infrared Spectrum . . . . .	29
3.2 Face Image Reconstruction . . . . .	30
3.3 Attribute Recognition . . . . .	32

<b>4</b>	<b>Visual Understanding from Thermal Infrared Image</b>	<b>35</b>
4.1	Background	36
4.2	Learning-based Reconstruction Framework	36
4.2.1	Global Reconstruction Step	38
4.2.1.1	Training Phase	38
4.2.1.2	Reconstruction Phase	40
4.2.2	Local Refinement Step	43
4.2.2.1	Training Phase	43
4.2.2.2	Reconstruction Phase	45
4.3	Experiments and Analysis	47
4.3.1	Dataset	47
4.3.2	Experimental Setup	48
4.3.3	Comparative Methods	50
4.3.4	Results and Analysis	51
4.4	Discussion	56
4.4.1	Finding the Optimal Set of Parameters	56
4.4.2	Face Image Reconstruction of Unknown Persons	59
4.4.2.1	Performance Evaluation	59
4.4.2.2	Various Sizes of Training Dataset	63
4.5	Summary	67
<b>5</b>	<b>Semantic Understanding from Thermal Infrared Image</b>	<b>69</b>
5.1	Background	70
5.2	Robust Principal Component Analysis (RPCA)	70
5.2.1	RPCA via Principal Component Pursuit (PCP)	71
5.2.2	RPCA via Principal Component Pursuit (PCP) with Probability Map	72
5.3	Recognition Framework	74
5.3.1	Recognition Framework with RPCA via PCP	76
5.3.2	Recognition Framework with RPCA via PM-PCP	78
5.3.2.1	Creation of Probability Map	78
5.3.2.2	Decomposition and Recognition Phases	79
5.4	Experiments and Analysis	81
5.4.1	Dataset	82
5.4.2	Experimental Setup	84
5.4.3	Comparative Methods	84
5.4.4	Results and Analysis	87
5.5	Summary	91
<b>6</b>	<b>Conclusion</b>	<b>95</b>
6.1	Thesis Summary	95
6.2	Future Work	97

---

6.3 Closing Remarks . . . . .	99
<b>Bibliography</b>	<b>101</b>
<b>List of Publications</b>	<b>119</b>



# List of Figures

1.1	Electromagnetic spectrum at a glance. UV stands for Ultraviolet. . . .	2
1.2	Image examples in different spectra. . . . .	3
1.3	Application examples where human intervention is necessary. . . .	6
1.4	Two main research questions in this thesis. . . . .	7
1.5	Image examples of a person with wearable attributes in different spectra. . . . .	9
1.6	Overview of the reconstruction. . . . .	11
1.7	Overview of the recognition. . . . .	14
1.8	Dataset example with five images where attributes are the minority. .	15
1.9	Structure of the thesis. . . . .	17
2.1	Examples of images taken in (from left to right) visible, near-infrared and thermal infrared spectra. Image courtesy of Espinosa-Duro et al. [34]. . . . .	24
3.1	Overview of the related work. . . . .	28
4.1	Process flow of the proposed reconstruction method. . . . .	37
4.2	Overview of the Global Reconstruction step. . . . .	38
4.3	Overview of the LLE method. . . . .	41
4.4	Overview of the Local Refinement step. . . . .	43
4.5	Examples of patches taken from a face image. . . . .	44
4.6	Evaluation results of PSNR. . . . .	52
4.7	Evaluation results of SSIM. . . . .	52
4.8	Evaluation results of Face Recognition. . . . .	53
4.9	Examples of reconstructed images by various methods. Each row shows an individual person and the columns contain: (A) Ground-truth images; (B) Reconstructed images of the proposed method; (C) Reconstructed images of Holistic LLE (Locally Linear Embedding); (D) Reconstructed images of Patch-Based LLE; (E) Reconstructed images of Patch-Based 1-NN (Nearest Neighbor); (F) Reconstructed images of Patch-Based $k$ -NN. . . . .	53
4.10	Heat map representation of the confusion matrix of the recognition evaluation. It goes from dark blue to dark red, where the representations of higher values are close to dark red. . . . .	54

4.11	Visual examples in various steps of the proposed method. Each row shows an individual person and the columns contain: (A) Thermal infrared input images; (B) Globally reconstructed images; (C) Residual images (Normalized to 0–1); (D) Fully reconstructed images; (E) Ground-truth images. . . . .	55
4.12	Evaluation results of PSNR. . . . .	57
4.13	Evaluation results of SSIM. . . . .	58
4.14	Evaluation results of Face Recognition. . . . .	58
4.15	Evaluation results of PSNR. . . . .	61
4.16	Evaluation results of SSIM. . . . .	61
4.17	Evaluation results of Face Recognition. . . . .	62
4.18	Examples of reconstructed images by various methods of unknown persons. Each row shows an individual person and the columns contain: (A) Ground-truth images; (B) Reconstructed images of the proposed method; (C) Reconstructed images of Holistic LLE (Locally Linear Embedding); (D) Reconstructed images of Patch-Based LLE; (E) Reconstructed images of Patch-Based 1-NN (Nearest Neighbor); (F) Reconstructed images of Patch-Based $k$ -NN. . . . .	62
4.19	Evaluation results of PSNR. The PSNR was linearly extrapolated (red dashed line). . . . .	65
4.20	Evaluation results of SSIM. The SSIM was linearly extrapolated (red dashed line). . . . .	65
4.21	Evaluation results of Face Recognition. The recognition rate was linearly extrapolated (red dashed line). . . . .	66
4.22	Examples of reconstructed images with various numbers of training data. Each row shows an individual person and the columns contain: (A) Ground-truth images; (B) Reconstructed images from 30 subjects' training data; (C) Reconstructed images from 60 subjects' training data; (D) Reconstructed images from 90 subjects' training data; (E) Reconstructed images from 120 subjects' training data; (F) Reconstructed images from 150 subjects' training data. . . . .	66
5.1	Usage of RPCA for decomposition. . . . .	75
5.2	Process flow of the recognition framework by RPCA via PCP. . . . .	76
5.3	Process flow of the recognition framework by RPCA via PM-PCP. . . . .	78
5.4	Example of thermal image and various Probability Maps (PMs). Each image represents: (A) Thermal image with <i>hat</i> , <i>glasses</i> , and <i>mask</i> attributes; (B) <i>Glasses</i> PM; (C) <i>Mask</i> PM; (D) <i>Shoulder bag</i> PM; (E) <i>Backpack</i> PM . . . . .	80
5.5	Image examples from the dataset. . . . .	83
5.6	Evaluation results by F-score per attribute. . . . .	88
5.7	Evaluation results by F-score per region. . . . .	89

- 
- 5.8 Examples of decomposition results by various methods. Images represent: (A) Original thermal infrared; (B) Results of Conventional Average; (C) Results of Conventional PCA; (D) Results of RPCA via PCP; (E) Results of Post-PCP PM; (F) Results of RPCA via PM-PCP. The decomposition is made for (from top to bottom): *glasses*, *mask*, *shoulder bag*, and *backpack* attributes. For visualization purpose, the minimum and maximum pixel values for each image were normalized to 0 and 1, respectively. . . . . 90
- 5.9 Examples of decomposition results by methods that use Probability Map in the case of negative input images. Images represent: (A) Original thermal infrared; (B) Results of Post-PCP PM; (C) Results of RPCA via PM-PCP. The decomposition is made for (from top to bottom): *backpack*, *shoulder bag*, *glasses*, and *mask* attributes. For visualization purpose, the minimum and maximum pixel values for each image were normalized to 0 and 1, respectively. Since the images do not contain the target attributes, pixels of the decomposed images are close to black. . . . . 92



# List of Tables

2.1	Divisions in the electromagnetic spectrum. . . . .	20
2.2	Spectral bands division defined by International Organization for Standardization (ISO) [28]. . . . .	22
2.3	Divisions in the infrared spectrum [30]. . . . .	22
2.4	Divisions in the infrared spectrum according to International Organization for Standardization (ISO) [28]. . . . .	22
4.1	Comparison of the methods performed in the experiment. . . . .	50
4.2	Comparison of the proposed method with the holistic method and various patch-based methods (Number of neighbors: 5, Patch size: $9 \times 9$ pixels). SE stands for the standard error, and is represented in the bracket. . . . .	51
4.3	Evaluation of various patch sizes with various numbers of neighbors. SE stands for the standard error, and is represented in the bracket. . . . .	57
4.4	Comparison of the proposed method with the holistic method and various patch-based methods (Number of neighbors: 5, Patch size: $9 \times 9$ pixels). SE stands for the standard error, and is represented in the bracket. . . . .	60
4.5	Evaluation of different numbers of training data (Number of neighbors: 5, Patch size: $9 \times 9$ pixels). SE stands for the standard error, and is represented in the bracket. . . . .	64
5.1	Distribution of the seven wearable attributes in the dataset. Note that there are images that contain more than one wearable attributes. . . . .	83
5.2	Comparison of the methods performed in the experiment. . . . .	85
5.3	Comparison of the proposed methods with other methods in F-Score. . . . .	86



# Abbreviations

<b>CCA</b>	<b>Canonical Correlation Analysis</b>
<b>DCCA</b>	<b>Deep Canonical Correlation Analysis</b>
<b>FIR</b>	<b>Far InfraRed</b>
<b>HOG</b>	<b>Histogram of Oriented Gradients</b>
<b>HSV</b>	<b>Hue, Saturation, Value</b>
<b>ISO</b>	<b>International Organization for Standardization</b>
<b>IR</b>	<b>InfraRed</b>
<b>KCCA</b>	<b>Kernel Canonical Correlation Analysis</b>
<b>LBP</b>	<b>Local Binary Patterns</b>
<b>LLE</b>	<b>Locally Linear Embedding</b>
<b>LWIR</b>	<b>Long-Wavelength InfraRed</b>
<b>MIR</b>	<b>Mid InfraRed</b>
<b>MRF</b>	<b>Markov Random Fields</b>
<b>MWIR</b>	<b>Mid-Wavelength InfraRed</b>
<b>NIR</b>	<b>Near InfraRed</b>
<b>NN</b>	<b>Nearest Neighbor(s)</b>
<b>PCA</b>	<b>Principal Component Analysis</b>
<b>PCP</b>	<b>Principal Component Pursuit</b>
<b>PM</b>	<b>Probability Map</b>
<b>PM-PCP</b>	<b>Probability Map - Principal Component Pursuit</b>
<b>PSNR</b>	<b>Peak Signal to Noise Ratio</b>
<b>RFR</b>	<b>Radio-Frequency Radiation</b>
<b>RGB</b>	<b>Red, Green, Blue</b>

<b>RPCA</b>	<b>Robust Principal Component Analysis</b>
<b>RR</b>	<b>Recognition Rate</b>
<b>SE</b>	<b>Standard Error</b>
<b>SSIM</b>	<b>Structural SIMilarity</b>
<b>SWIR</b>	<b>Short-Wavelength InfraRed</b>
<b>UV</b>	<b>Ultra Violet</b>

*Dedicated to my wife, Mia.*



# Chapter 1

## Introduction

This chapter introduces the research presented in the thesis. It starts with the general background in Section 1.1, then continues into the details of the theme which is understanding human images in the thermal infrared spectrum in Section 1.2. Research topics affiliated with the theme are explained in Section 1.3, and lastly, the description of the general structure of the thesis follows in Section 1.4.

### 1.1 Background

In this day and age, security surveillance systems can be seen almost everywhere. Although they are generally found in commercial and public buildings, it is increasingly common to find them in residential buildings as well. These surveillance systems play a critical role in security as means for crime prevention, by deterring potential criminals; as well as for monitoring and crime investigations [1, 2].

There are many factors to consider when setting up and managing a surveillance system. For example, a need to consider the locations to be covered by the cameras, the security level to be enforced, and the type and quality of cameras desired. It is also important to note that security surveillance systems work continuously, every

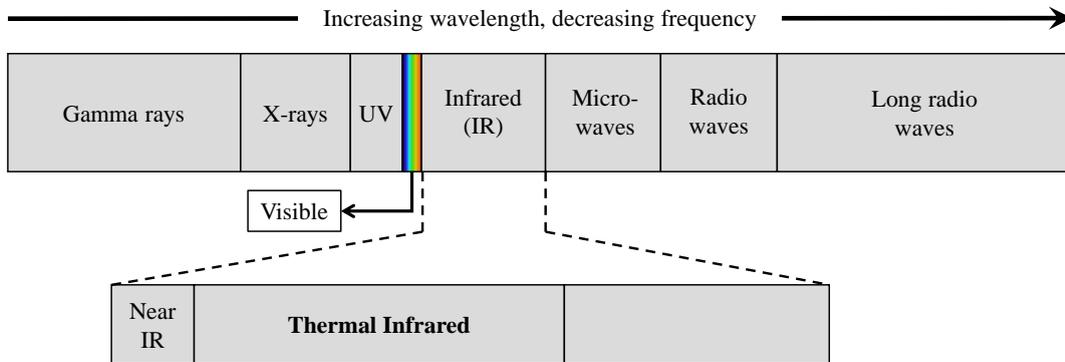


FIGURE 1.1: Electromagnetic spectrum at a glance. UV stands for Ultraviolet.

single day, through day and night. Therefore, illumination and weather conditions also need to be considered.

Locations covered by surveillance cameras can be divided into two categories: indoors and outdoors. Indoor surveillance systems are relatively easier to handle because the lighting can be controlled. However, the lighting must be able to provide the best illumination condition possible. For example, it is preferable to have ideal illumination in an area close to one of the cameras and forsake some lighting where the area is relatively far away. There are not many other external factors that can compromise the quality of indoor images. That being said, a less than ideal illumination condition is still present even indoors due to the possibility of light sources being turned off during the night, and non-uniform lighting in the scene [3–5].

On the other hand, in outdoor environments, the illumination conditions vary according to the time of day. During the night, when there is no sunlight, humans must rely on artificial light sources. Unless additional efforts are made to brighten all of the areas monitored by surveillance cameras, achieving an ideal illumination condition is a massive task. Lastly, weather conditions such as fog or rain can also affect the capability to capture an ideal image or video.

There are many types of cameras that can be used in surveillance systems, which may be categorized by the spectrum they are working in. Figure 1.1 shows the electromagnetic spectrum in its simplest representation. The most common camera works in the visible spectrum that ranges from about 400 to 700 nm in the electromagnetic

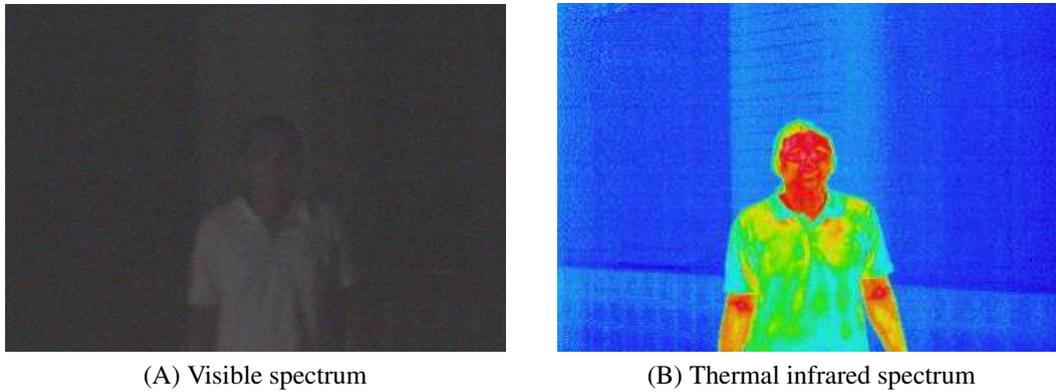


FIGURE 1.2: Image examples in different spectra.

spectrum, in which the human visual system also operates. A general market consumer usually purchases these cameras, which this thesis will refer to as “normal cameras” or simply “cameras” from hereon. These cameras are relatively inexpensive and can be easily purchased. A basic security surveillance system employs normal cameras to cover the desired areas, which is usually sufficient. However, these cameras face a big problem during nighttime or in dark areas. In less-than-ideal illumination conditions, whether indoors or outdoors, the security surveillance system cannot stop performing its duty. Some solutions to this problem are providing light sources on affected areas or limiting entry to certain places.

As an alternative to these normal cameras, thermal cameras can also be utilized for a surveillance system. Thermal cameras work in thermal infrared spectrum, a subdivision of the infrared spectrum, and do not rely on illumination to produce images. Rather than capturing visible light, these cameras capture infrared radiation whose intensity depends on the temperature of the object [6, 7]. This capturing process is called thermal imaging or thermography. This characteristic makes a thermal camera a better option over normal cameras for surveillance in nighttime and/or poorly lit areas. Figure 1.2 shows an example of images in both visible and thermal infrared spectra taken simultaneously during the night. Although there is a light source nearby, as seen in Figure 1.2(A), it is hard to discern any feature from the person. Meanwhile, the thermal infrared version of the image shows the presence of a human clearly.

Other than these two camera types, there are also cameras available for surveillance that work in near-infrared spectrum, another sub-division of the infrared spectrum. Near-infrared cameras also allow humans to see the surrounding environment in the dark by using active infrared lighting, which is the main difference with the thermal infrared cameras. The cameras then capture the infrared wave reflected from objects to produce images. Images captured in the near-infrared spectrum resemble ones that are captured by normal cameras, but with slight differences. For example, in a face image, the person's eyes look completely dark and the skin lacks texture. Further discussion on both the near-infrared and the thermal infrared spectra is provided in [Chapter 2](#).

Currently, the most common implementation of a security surveillance system is to have a dedicated person, usually a security guard, to observe various monitors from multiple camera feeds. This is a taxing assignment, and a normal human cannot stay looking at these camera feeds without a break. On top of being very labor intensive, it is also prone to human error. Consequently, the surveillance system cannot be used reliably for continuous tasks such as face recognition and human detection. In the condition where nobody is watching the camera feeds, at best, the surveillance system can only be used to conduct investigations of crimes after they have occurred through recording.

Automating some tasks of the surveillance system can be very beneficial. While a human observer may still be necessary, the assignment will not be as taxing. For example, an automated human detection system in highly restricted areas, where the presence of humans in these areas equals to a security risk. A quick detection guarantee sufficient time for an appropriate action, and by automating this process, it does not rely on a dedicated person watching the camera feeds continuously [8, 9].

In some cases, human intervention is necessary even with automations. Take an automated face recognition system used as a security check at the entrance to a building. When the system recognizes a person suspected to be wanted by the police, it can

alert a security guard to confirm this hypothesis. If the system did not make a mistake, the security guard then can decide on the steps taken next; contact the police or apprehend the person directly. Figure 1.3(A) visualizes this example.

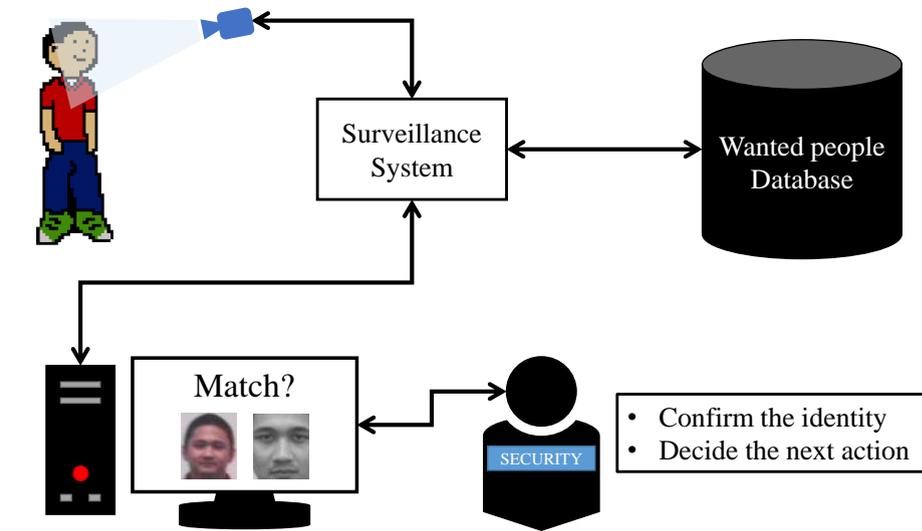
Another example is when an automated system fails to recognize an individual due to the usage of certain objects. In areas where the identity of a person is important, such as a bank or a military outpost, this is a security risk. Human intervention might be necessary to take further action in order to identify the individual. This case is visualized in Figure 1.3(B).

Depending on the desired security level, more automation can be introduced into the surveillance system. In a situation where higher level of security is imposed, the surveillance system benefits from the existence of additional and appropriate automated systems, including but not limited to the previously mentioned face recognition and human detection systems.

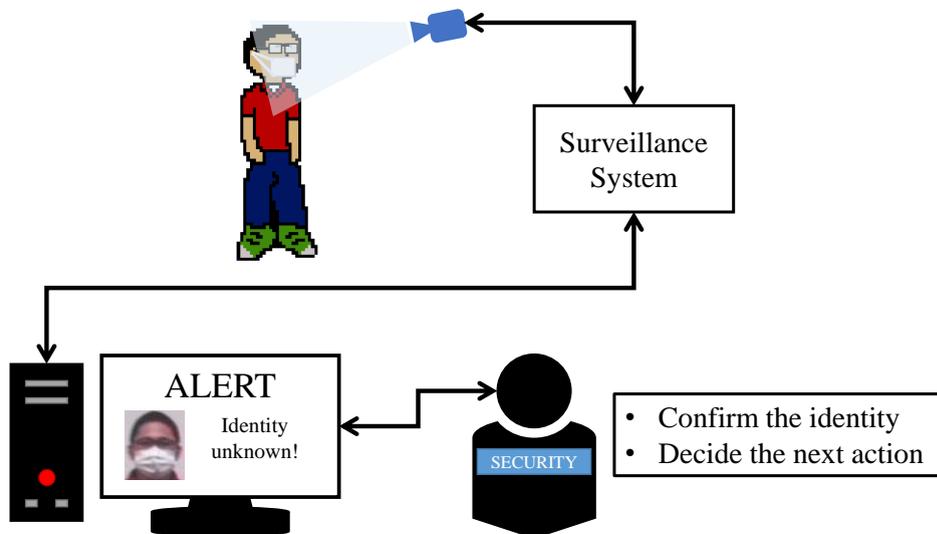
Although thermal infrared cameras have some advantages over normal cameras, there are still some drawbacks. Generally, humans will encounter great difficulty in observing thermal infrared images, and cannot easily identify a person captured in them. This is disadvantageous in surveillance, since human participation is necessary in certain applications mentioned previously. In these applications, the capability to recognize humans captured by the surveillance camera is quite important. However, many features of a face available in the visible spectrum are not necessarily usable in the thermal infrared spectrum, making the task of identifying people captured in the thermal infrared spectrum a tremendous challenge. This brings us to the main theme of this thesis, namely “Understanding thermal infrared images.”

## **1.2 Understanding Thermal Infrared Images**

There are two parts in this theme; “thermal infrared” and “image understanding”. Thermal infrared is a name of a spectrum which is a sub-division of the infrared spectrum, as explained briefly in the previous section. Image understanding, on the



(A) Example where the surveillance system finds a person suspected to be wanted by police.



(B) Example where the surveillance system fails to identify a person due to certain objects.

FIGURE 1.3: Application examples where human intervention is necessary.

other hand, is a familiar term in the field of computer vision. While there is no agreed definition of this term, it can generally be explained as *inferring facts on what is in the image based on its elements*. This may include the usage of statistics, geometry, and other scientific subjects to do so. Therefore, the combination of these two parts means that the image that this thesis investigates to gain understanding is taken in the thermal infrared spectrum.

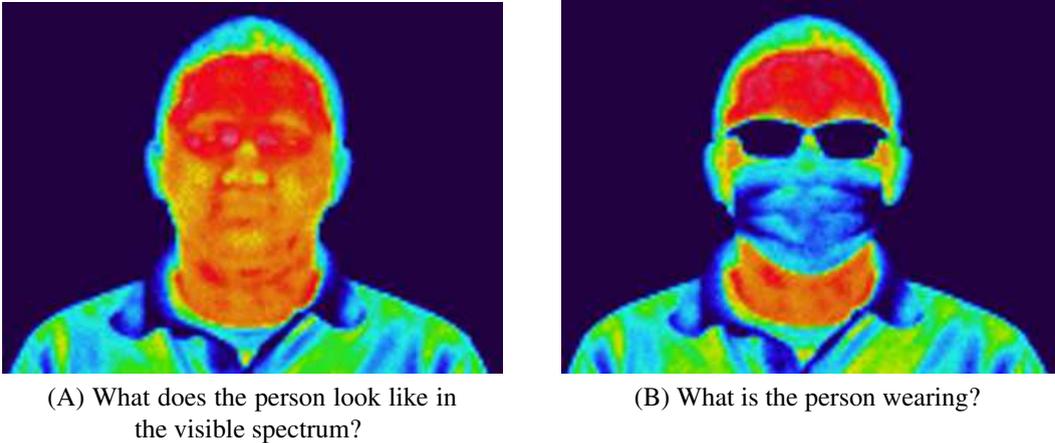


FIGURE 1.4: Two main research questions in this thesis.

Understanding thermal infrared image in general is an extensive task. However, in regards to the surveillance, the most common target of observation is humans. Therefore, the focal point of the research in this thesis is understanding human images captured in the thermal infrared spectrum. However, since the human visual system does not work in the thermal infrared spectrum, it is difficult to understand human images captured in this spectrum. The main cause of difficulty in performing human identification is due to the lack of understanding of human images in the thermal infrared spectrum. In other words, we do not understand how humans look and are represented in this spectrum. Rather than forcing humans to understand the human images, a better option would be to leave the task of understanding to a system. Then, the system can be used to help people understand thermal infrared images in various applications.

From human images captured in the thermal infrared spectrum, there may be multiple research questions. Two of them can be seen in Figure 1.4, which are “What does the person look like in the visible spectrum?” and “What is the person wearing?” These two questions are the foci of the research presented in this thesis, and answering them would help humans understand more of thermal infrared images.

### **1.2.1 Research Question 1: What does the person look like in the visible spectrum?**

In general, humans find it difficult to recognize/identify who is the person in a thermal infrared image. This is because features available in the visible spectrum might not be useful in the thermal infrared spectrum. This is the main reason why this question is being asked.

Nowadays, multiple automated surveillance systems have been proposed [10, 11], despite the fact that there are still some issues in surveillance [12, 13]. In the visible domain, humans can easily confirm the hypothesis made by the automated system. One example was given in Section 1.1 where the system recognizes a person suspected to be wanted by the police, and humans are required to confirm this hypothesis. However, in the case where thermal cameras are used, humans are less likely to be able to neither confirm nor deny this since we lack the understanding necessary to do so.

Therefore, in order to answer this question, the human image needs to be reconstructed from the thermal infrared spectrum to the visible spectrum. Doing so provides visualization to humans, which is very beneficial in the case of using thermal infrared cameras for surveillance. To be able to achieve this, understanding the correlation between face images in both spectra is necessary.

### **1.2.2 Research Question 2: What is the person wearing?**

In daily life, it is not uncommon to see people wearing glasses, accessories and other vanity items on top of basic clothing. However, the existence of these objects on the face is an occlusion, hindering the identification process. This fact applies not only to systems, but also humans. In surveillance, this is not an ideal situation and as a matter of fact, it is a security risk, especially in areas where the identity of a person is important such as a bank or a military outpost. Additionally, what people can bring

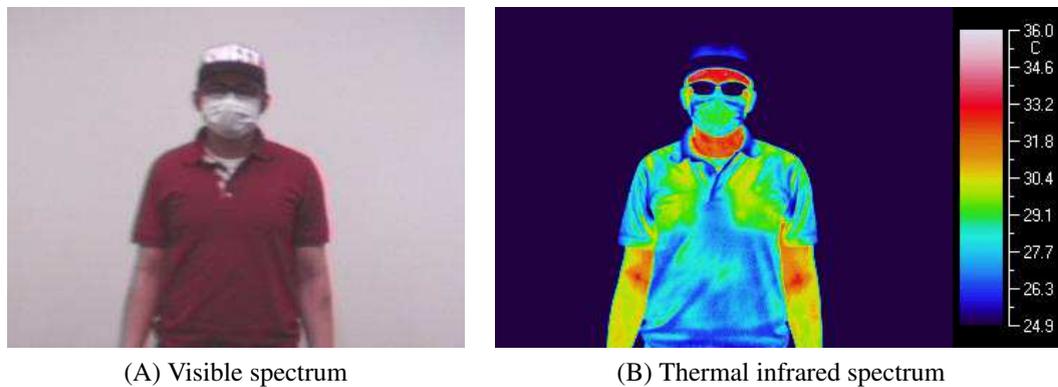


FIGURE 1.5: Image examples of a person with wearable attributes in different spectra.

into these places are often regulated. Therefore, answering this research question by recognizing what attributes people are wearing is important.

The term “attribute” can be defined as a trait or an element of an object. Humans can have a lot of attributes affiliated with them, encompassing various aspects. Age, race, gender, skin tone, hair type, hair color, are just a few examples which are directly correlated to the human body. Other examples are related to the numerous objects that can be worn. Aside from the many variations of basic clothing, there are abundant amount of accessories, bags, and headgears that could be worn by humans. In this research, these worn objects are referred to as “wearable attributes.”

In surveillance systems, human identification requires analysis and processing of the image of the related person. However, the presence of wearable attributes in the image may hinder this process. In these cases, prior to the identification, there is a need to recognize the wearable attributes present in the image. In the visible spectrum, recognition of these wearable attributes can be used to describe people [1, 14]. In the thermal infrared spectrum however, the temperature of an object is an integral part of the imaging process, hence these attributes are seen differently. Some of these attributes cannot be used and may even hinder the identification process. In this case, it is a better decision to isolate them because they can also be a security risk. Figure 1.5 shows an example of images of a person with wearable attributes in both visible and thermal infrared spectra.

There are two considerable problems that need to be addressed in order to recognize these wearable attributes. The first problem is related to the relatively small size of the wearable attributes when compared to the size of the human body. Since they are small, the difficulty of the recognition rises. The second problem is the existence of variations in both the human body and the wearable attribute themselves, which means that the intra-class variation is high. Consequently, recognizing these wearable attributes is a challenging task.

### **1.3 Research Topics**

By expanding on the concept of image understanding, understanding human images could be rephrased as *inferring facts about the person(s) in the image based on its elements*. In this thesis, two approaches were chosen to understand these human images; visually and semantically. Respectively, both “visual understanding” and “semantic understanding” can be obtained from human images. Analysis of elements of the person from the appearance and what is seen visually in the image is performed to gain visual understanding. As for obtaining semantic understanding, facts can be inferred from the existence or the absence of human elements in the image and describe the image with it.

To shed some light into these two concepts of understanding in this thesis, two research topics will be tackled to investigate each of these concepts separately. The first research topic is related to the investigation of the visual understanding of human images in the thermal infrared spectrum, which is performed to answer the first research question. The second research topic is related to the investigation of the semantic understanding of human images in the thermal spectrum, which is performed to answer the second research question.

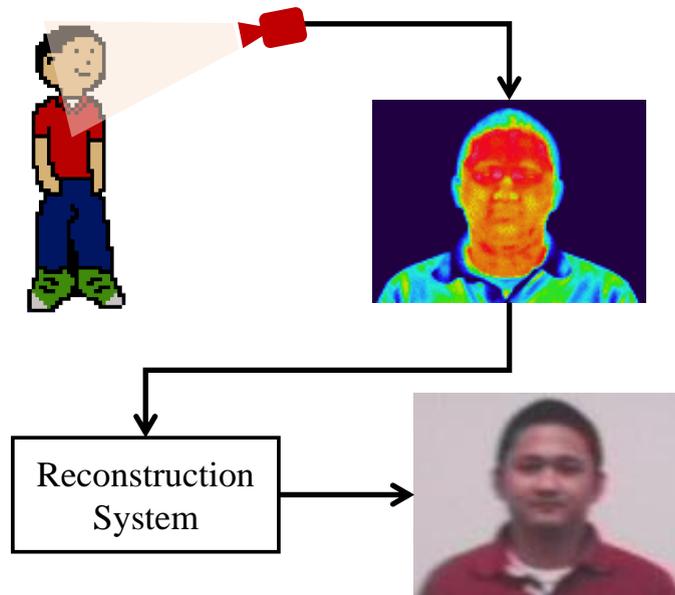


FIGURE 1.6: Overview of the reconstruction.

### 1.3.1 Topic 1: Visual Understanding from Thermal Infrared Image

In order to answer the first research question, which is “What does the person look like in the visible spectrum?”, this research topic attempts to reconstruct a face image from the thermal infrared spectrum to the visible spectrum. Figure 1.6 shows the overview of the reconstruction.

It has been mentioned previously that features of a face available in the visible spectrum might not be usable in the thermal infrared spectrum. However, regardless of the spectrum the image was taken in, the biometric information of a face is unique to the person. Based on this reasoning, this research assumed that a correlation can be inferred between the biometric information obtained from face images captured in both the thermal infrared and the visible spectra. The correlation is then used to perform the reconstruction.

From the various ways to study the correlation between the two modalities, this research chose to present thermal infrared and visible images into their low-dimensional subspace and perform the study on these low-rank representations. Based

on the transformation process, the low-rank representations will have various levels of information compacted from the original image. Since the correlation is derived from these low-rank representations, the relationship between the thermal infrared and the visible images can presumably be studied from a deeper level.

The proposed method includes the study of the correlation between the face images in thermal infrared and visible spectra, which will be used for the reconstruction. The proposed method is separated into two processes; the first is the learning process and the second is the reconstruction process.

The learning process comprises of representing the face images to their low-rank versions and the study of the correlation between the thermal infrared and visual versions. To obtain the low-rank versions of the face images, Principal Component Analysis (PCA) [15] is utilized. With PCA, the face images are represented in the lower-rank version of themselves on the eigenspace. It is important to note that the eigenspaces of the thermal infrared and the visible spectra are not the same, meaning each spectrum has its own eigenspace. After the low-rank versions of the face images are available on the eigenspace, Canonical Correlation Analysis (CCA) [16] is employed to learn the correlation between the thermal infrared and the visible face data. The CCA process projects low-rank face data from both the thermal infrared and the visible spectra onto a coherent space, where the correlation between the face data is maximized. On this coherent space, the correlation between the thermal infrared and the visible spectra can be exploited to achieve cross-modality reconstruction.

The reconstruction process is bidirectional, where the projection of a new face image can be performed from either the thermal infrared or the visible spectra. This research focuses on the face image reconstruction from the thermal infrared spectrum to the visible spectrum for the intended application in the visualization process in a surveillance system. Since the correlation between face images in both the thermal infrared and the visible spectra is studied on the eigenspace, and accordingly, on the coherent space where the maximized correlation can be derived. Therefore, it is necessary to project the new face image onto the eigenspace and the coherent space,

respectively. The reconstruction is performed on this coherent space by utilizing the Locally Linear Embedding (LLE) [17] method. The LLE is used to reconstruct the visible spectrum face data from the thermal infrared counterpart. After that, the data is projected back onto the image space and is produced in the form of an image.

In regards to the reconstruction, there is another important point that is also considered in the proposed method. Since the learning process includes a transformation of the data into its lower-dimensional space, details of the face are lost during the whole reconstruction procedure. Therefore, the proposed method employs a two-step reconstruction approach, which denotes the scale of the reconstruction performed on the image. The first step is a reconstruction on the whole face image, where the whole face image is created. Conversion to low-rank that causes the loss of details mentioned previously, is performed in this step. To reintroduce the details, it is necessary to perform the second step of reconstruction. Using the face image reconstructed in the first step as a foundation, the second step of reconstruction is performed. The second step involves reconstruction of small patches of the face image. By performing the reconstruction on small patches and without performing low-rank conversion, details of the face are preserved during the process. With these two steps, the proposed method takes advantage of both the global face structure and local features to perform the reconstruction. The details of the proposed framework is explained in Chapter 3.

### **1.3.2 Topic 2: Semantic Understanding from Thermal Infrared Image**

In order to answer the second research question, which is “What is the person wearing?”, this research topic attempts to recognize wearable attributes in the human image. Figure 1.7 shows the overview of the recognition.

To recognize these wearable attributes based on what is available in the image, a decomposition approach is proposed. One of the characteristics of wearable attributes

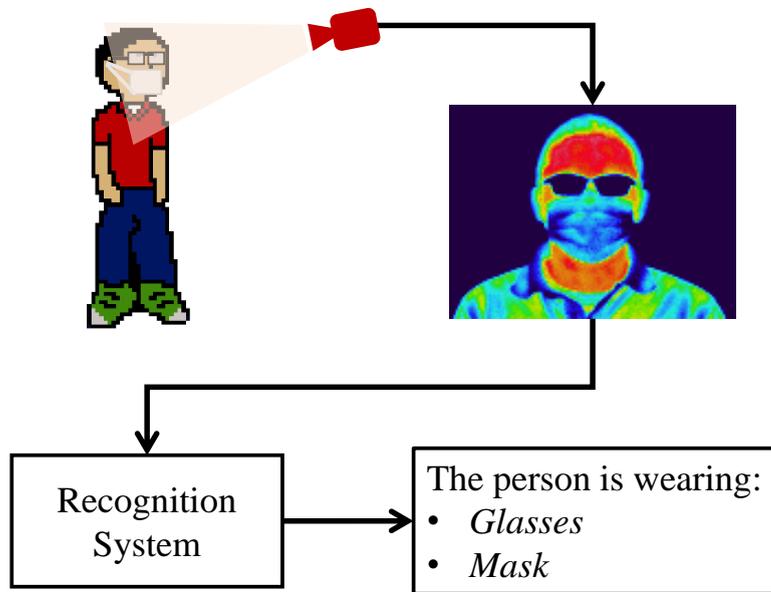


FIGURE 1.7: Overview of the recognition.

is that they usually have a lower temperature compared to those of a human body. This characteristic can be exploited by a decomposition approach. By decomposing the human image to multiple components, the components can then be analyzed. This research proposes the usage of a low-rank representation of the image to obtain other components of the image to achieve the decomposition. With this information, a conclusion can then be drawn on the existence of wearable attributes in the image.

The proposed method employs a decomposition process that could separate the image into multiple components, one of which contains the wearable attributes. Since the attributes are all that is needed, the other elements are unnecessary. Therefore, this research proposes the idea of majority and minority to achieve the decomposition by utilizing the low-rank representations. To realize the idea, there is a condition that needs to be fulfilled. The condition is that in a group of observations, the similarity between the observations needs to be high.

In order to convey the idea in a simpler fashion, let us assume a collection of human images in the thermal infrared spectrum. Most of these people are shown without wearable attributes, serving as the majority. In this case, the existence of a person with a wearable attribute in some images is evaluated as the minority. Additionally,

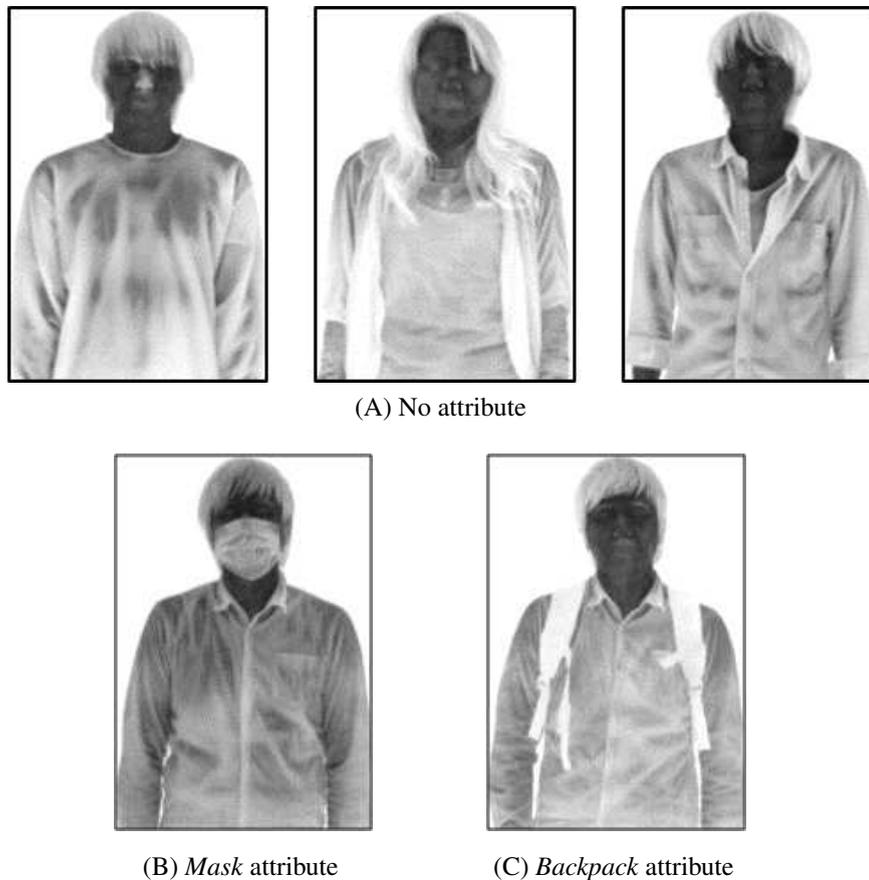


FIGURE 1.8: Dataset example with five images where attributes are the minority.

the small size of the attributes, which is one of the problems mentioned in Section 1.2.2, can also be considered as minority and the rest of the human parts are the majority. Under this assumption, the attributes can be extracted by means of decomposition. To ensure this condition, the proposed method exercises control over the data prior to the decomposition. Figure 1.8 depicts an example of this idea.

The decomposition in this research is performed by utilizing Robust Principal Component Analysis (RPCA) [18], a modification to the popular PCA. Compared to PCA, RPCA is robust to corrupted or noisy observations in the data. RPCA can decompose noisy images into two components; a component which contains only noise and another component that depicts the original image without noise. The aforementioned decomposition idea considers the wearable attributes as a minority,

which in the case of RPCA, as noise. Therefore, the usage of RPCA for decomposition is beneficial for our purpose, as the wearable attributes can be extracted into a separate component.

The recognition process takes part after the decomposition. In addition, since the wearable attributes are isolated into individual components, any classifier can be used to accomplish the recognition.

In this research, the proposed method utilizes the RPCA for the decomposition. First, the RPCA is used as is for the decomposition. Afterwards, this method is improved with the usage of prior knowledge. The prior knowledge used is the region where the wearable attributes are likely to be present. In this work, this prior knowledge is named as the Probability Map (PM). The PM shows the likelihood of an attribute being present in the region of an image which is deduced from the available annotated data. The frameworks of the proposed RPCA methods are described in Chapter 4.

## 1.4 Structure of the Thesis

This thesis is divided into five chapters as seen in Figure 1.9. Chapter 1 covered the background of the thesis and introduced two research topics in this thesis and the approaches taken to propose the solution for each topic. Chapter 2 explains the thermal infrared spectrum in detail, while Chapter 3 describes various related researches that have been conducted. Previous works related to the two research topics are also included in this chapter. Chapter 4 goes into the detail about the first research topic; Visual understanding of a face image via reconstruction from the thermal infrared to the visible spectra. Chapter 5 describes the second topic; Semantic understanding of a human image via recognition of wearable attributes in thermal infrared images. Both of these chapters include a full explanation of the proposed methods, the experiments conducted and their respective analyses, and closed by their own summaries. Finally, Chapter 6 concludes this thesis with the summary, future works, and some closing remarks.

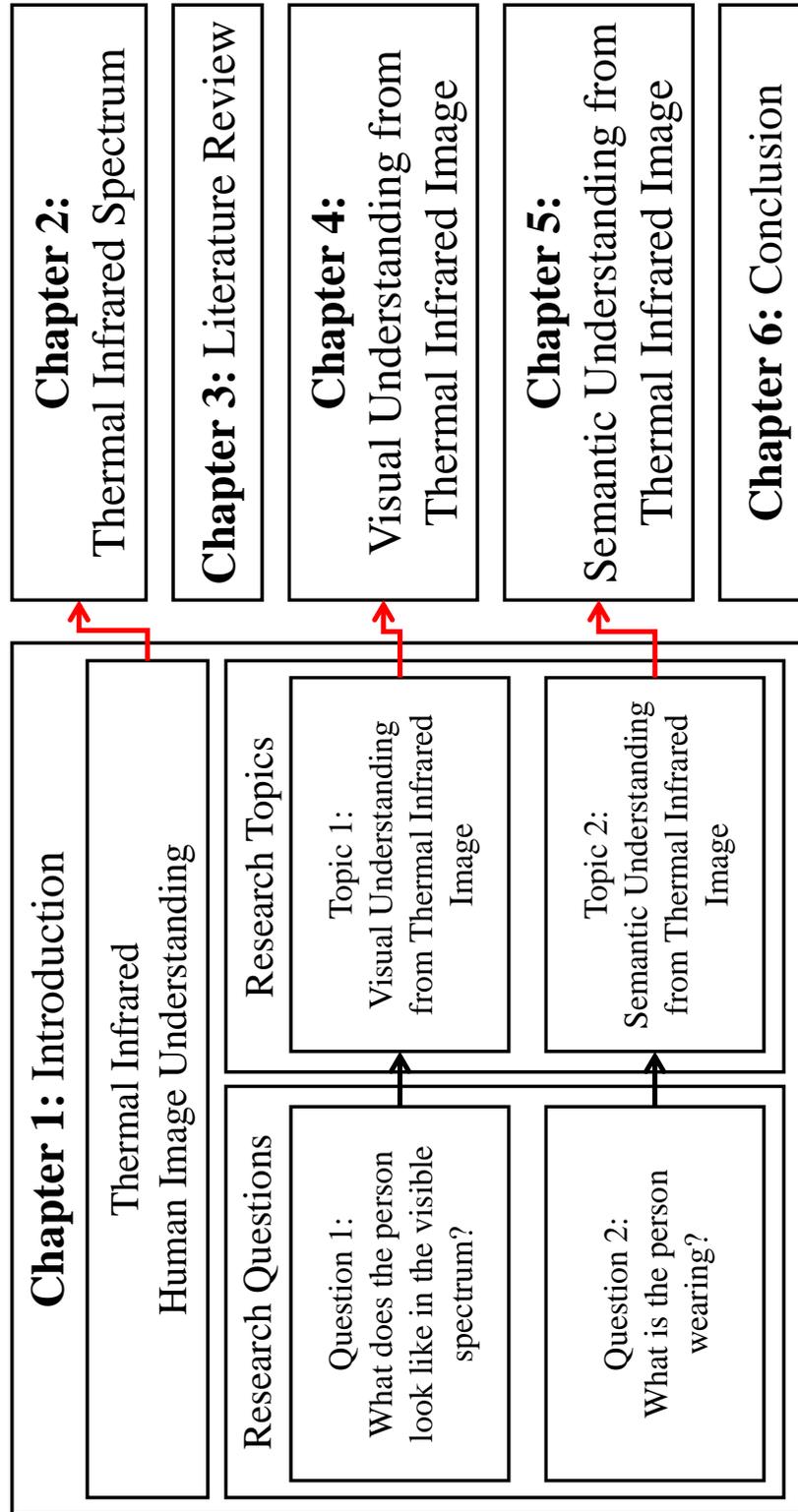


FIGURE 1.9: Structure of the thesis.



# Chapter 2

## Thermal Infrared Spectrum

This chapter introduces the spectrum of interest in the thesis; thermal infrared. It starts by introducing the broad electromagnetic spectrum in Section 2.1. Section 2.2 explains the infrared spectrum further, along with the thermal infrared spectrum.

### 2.1 Electromagnetic Spectrum

The electromagnetic spectrum refers to a collective band of electromagnetic radiation, extending from gamma rays to radio waves. The simplest representation of the electromagnetic spectrum is shown in Figure 1.1 in Chapter 1. This radiation, in the form of a wave, has a specific wavelength and frequency. One of the characteristics of a wave is that the wavelength and the frequency are inversely proportional to each other. Therefore, the longer the wavelength of a wave, the frequency of the said wave is lower. The electromagnetic spectrum is divided into multiple bands or spectra depending on the wavelength/frequency.

Starting from the shortest wavelength (thus highest frequency), various bands of the electromagnetic spectrum and their brief introductions are provided in the following list. Table 2.1 shows the division of the electromagnetic spectrum, including the wavelength and the frequency of each spectra [19].

TABLE 2.1: Divisions in the electromagnetic spectrum.

Division Name	Wavelength [m]	Frequency [Hz]
Gamma ray	$< 10 \times 10^{-12}$	$> 30 \times 10^{18}$
X-ray	$10 \times 10^{-12} - 10 \times 10^{-9}$	$30 \times 10^{15} - 30 \times 10^{18}$
Ultraviolet	$10 \times 10^{-9} - 400 \times 10^{-9}$	$790 \times 10^{12} - 30 \times 10^{15}$
Visible	$400 \times 10^{-9} - 700 \times 10^{-9}$	$430 \times 10^{12} - 790 \times 10^{12}$
Infrared	$700 \times 10^{-9} - 1 \times 10^{-3}$	$300 \times 10^9 - 430 \times 10^{12}$
Microwave	$1 \times 10^{-3} - 1 \times 10^0$	$300 \times 10^6 - 300 \times 10^9$
Radio wave	$1 \times 10^0 - 100 \times 10^3$	$3 \times 10^0 - 300 \times 10^6$

- **Gamma rays:** Also referred to as gamma radiation, denoted by the Greek letter  $\gamma$ . It has a very penetrating characteristic, discovered from the study of radioactive materials. It was discovered after alpha  $\alpha$  and beta  $\beta$  rays, hence its name [20]. The wavelength of gamma rays is generally less than 10 pm (picometers) or  $10^{-11}$  m.
- **X-rays:** The first published work on X-rays is by Röntgen [21] in 1896. He referred to this radiation by the name “X” to signify it as an unknown radiation. It has since been commonly referred to as either X-rays or Röntgen radiation, and often used in the medical imaging field. The wavelength of X-rays generally ranges from 0.01 to 10 nm.
- **Ultraviolet:** The word ultra originated from Latin, which means *beyond*, while violet is the color in the visible spectrum that has the shortest wavelength. Therefore, ultraviolet literally means “beyond violet.” The wavelength of ultraviolet generally ranges from 10 to 400 nm.
- **Visible light:** The electromagnetic radiation in this band is visible to humans and capturing this radiation is the basis of how the human vision works. There are different definitions on the range of wavelength that is considered as visible lights. One of the narrowest definition of the visible light is from 420 to 680 nm [22, 23], while the broadest definition of the visible light is from 380 to 800 nm [24]. Under artificial conditions, human eyes can also see electromagnetic radiation with wavelength from 310 up to 1,050 nm [25].

- **Infrared:** The word “infra” originated from Latin, which means *below*. Infrared radiation has a longer wavelength than that of visible light. In the visible spectrum, red is the color that has the longest wavelength, which means it has the lowest frequency. Infrared has a lower frequency than the red part of the visible spectrum, hence the name. The wavelengths that are categorized as infrared generally range from 0.7 to 1,000  $\mu\text{m}$ . Further explanation is provided in Section 2.2.
- **Microwaves:** The prefix “micro” in the name does not suggest it has a wavelength of a micrometer ( $10^{-6}$  m) which is denoted by  $\mu\text{m}$ . Instead, the wavelength for microwaves generally ranges from 1 to 1,000 mm [26, 27]. Arguably, the most well-known usage of this electromagnetic radiation is in microwave ovens.
- **Radio waves:** Also often referred to as Radio-Frequency Radiation (RFR), radio waves are usually considered to have microwave radiation as its subset and commonly described by its frequency [26]. Radio waves are widely used in various applications such as radio communication, broadcasting, and radar, to name a few. The wavelength of radio waves generally ranges from 1 to 100,000 km.

It is important to note that the boundaries of each band in the electromagnetic spectrum are not categorically rigid. For example, the aforementioned visible light has multiple definitions. However, the fact that both infrared and ultraviolet are located next to the visible spectrum does not change. Nevertheless, the International Organization for Standardization (ISO) published their spectral bands division for ultraviolet, visible light and infrared [28] as shown in Table 2.2.

## 2.2 Infrared Spectrum

Infrared radiation or simply InfraRed (IR) was discovered by Herschel [29] in 1800. It is located between the visible and the microwave bands in the electromagnetic

TABLE 2.2: Spectral bands division defined by International Organization for Standardization (ISO) [28].

Division Name	Wavelength [nm]
Ultraviolet	1–380
Visible	380–780
Infrared	780–10 <sup>6</sup>

TABLE 2.3: Divisions in the infrared spectrum [30].

	Division Name	Wavelength [nm]
Reflected infrared	Near-InfraRed (NIR)	750–1,400
	Short-Wavelength InfraRed (SWIR)	1,400–3,000
Thermal infrared	Mid-Wavelength InfraRed (MWIR)	3,000–8,000
	Long-Wavelength InfraRed (LWIR)	8,000–15,000
	Far InfraRed (FIR)	15,000–1,000,000

TABLE 2.4: Divisions in the infrared spectrum according to International Organization for Standardization (ISO) [28].

	Division Name	Wavelength [nm]
IR-A	Near InfraRed (NIR)	750–1,400
IR-B		1,400–3,000
IR-C	Mid InfraRed (MIR)	3,000–50,000
	Far InfraRed (FIR)	50,000–1,000,000

spectrum. While the wavelength typically ranges from 0.7 to 1,000  $\mu\text{m}$ , the starting value might differ depending on the definition of the visible spectrum.

Just like the electromagnetic spectrum and many of its divisions, the infrared spectrum can also be separated into sub-divisions. According to D’Amico et al. [30], the infrared spectrum is commonly separated into multiple sub-divisions as seen in Table 2.3. The Near-InfraRed (NIR) and the Short-Wavelength InfraRed (SWIR) combined is usually called reflected infrared and the Mid-Wavelength InfraRed (MWIR) and Long-Wavelength InfraRed (LWIR) combined is often referred to as thermal infrared. Another division is published by ISO [28], as shown in Table 2.4.

Although human eyes can normally only see the electromagnetic radiation in the visible spectrum, there are reports of people who are able to see infrared up to 1,050 nm [25, 31] in certain conditions.

Reiterating what has been mentioned in Chapter 1, thermal infrared cameras do not rely on illumination to observe the environment. These cameras capture infrared radiation of an object where the intensity captured depends on its temperature. This process is called thermal imaging or thermography.

In order to understand how temperature can be used to observe the environment, explanation of physical bodies and their relationship to temperature is necessary. Planck's law describes that a black body emits electromagnetic radiation that depends solely only on the body's temperature, independent of its shape and composition [32]. A black body is an idealized physical body which satisfies three conditions [33]; It is a perfect emitter, a diffuse emitter, and the best energy absorber. One of the characteristics, a perfect emitter, means that a black body emits more energy than any other surface. Meanwhile, real world objects can only emit a fraction of the radiation emitted by a black body. This emissivity also influences the radiation captured by the thermal infrared cameras. Therefore, the image captured by a thermal infrared camera is affected by the temperature of an object and its emissivity.

As mentioned previously, both MWIR and LWIR combined is often referred to as thermal infrared. Therefore, the thermal infrared spectrum includes the wavelengths from 3 to 15  $\mu\text{m}$ . However, thermal cameras usually capture radiation in the LWIR range (from 8 to 14  $\mu\text{m}$ ) to produce images. It is important to note that LWIR is also sometimes called far-infrared [30], and some researches refer to the thermal infrared as such.

There are many terms that can be used interchangeably in the thermal infrared subject, for example, thermal imaging can also be called thermography. The thermal infrared cameras can also be called thermographic cameras. These cameras can also be referred to as thermographs, which by definition are instruments that record temperature variations in an area or over a period of time. A graphic or visual record



FIGURE 2.1: Examples of images taken in (from left to right) visible, near-infrared and thermal infrared spectra. Image courtesy of Espinosa-Duro et al. [34].

produced by a thermograph is called a thermogram. Therefore, thermal images can also be called thermograms. For convenience, henceforth only the terms “thermal cameras” and “thermal images” will be used in this work.

Another relatively familiar and established sub-division of the infrared spectrum is the near-infrared (NIR). In the infrared spectrum, it covers the wavelength immediately beyond the visible spectrum. Night vision cameras are usually associated with these NIR cameras. However, night vision basically means the ability to see in low light conditions, where there is not much, if any, visible light. By this definition, night vision extends from the near-infrared all the way to the thermal infrared spectrum. Therefore, the idea that night vision is achieved by only using near-infrared cameras is a misconception. The main difference between near-infrared and thermal infrared for human images is that near-infrared captures human features that closely resemble those taken in the visible spectrum, while thermal infrared captures only the temperature. Examples of images taken in visible, near-infrared and thermal infrared spectra are provided in Figure 2.1.

Night vision cameras commonly operate by utilizing either image intensifiers, active illumination or thermal imaging. Image intensifiers are devices that can produce image by increasing the intensity of available visible light and/or converting non-visible light sources from near-infrared or short wave infrared to visible. Active illumination makes use of an active source of illumination in the near-infrared and/or short wave infrared to facilitate the image capturing process. In conclusion, thermal

imaging is not the same with the other night vision methods, which closely resembles photography in the visible spectrum.



# Chapter 3

## Literature Review

This chapter provides a literature review of the various researches on image processing that have been conducted in this spectrum, and works related to the two research topics which were introduced in Chapter 1.

Under the umbrella of human image processing, the overview of the related works can be seen in Figure 3.1. Section 3.1 introduces various image processing researches conducted in the thermal infrared spectrum that are not related directly to the two researches introduced in this thesis. The following sections provide research history to shed some light on the two research topics; Face image reconstruction from the thermal infrared spectrum to the visible spectrum in Section 3.2, followed by wearable attribute recognition in human images captured in the thermal infrared spectrum in Section 3.3. From another point of view, these topics can be considered as an image to image (conversion) research and an image to class/category (recognition) research.

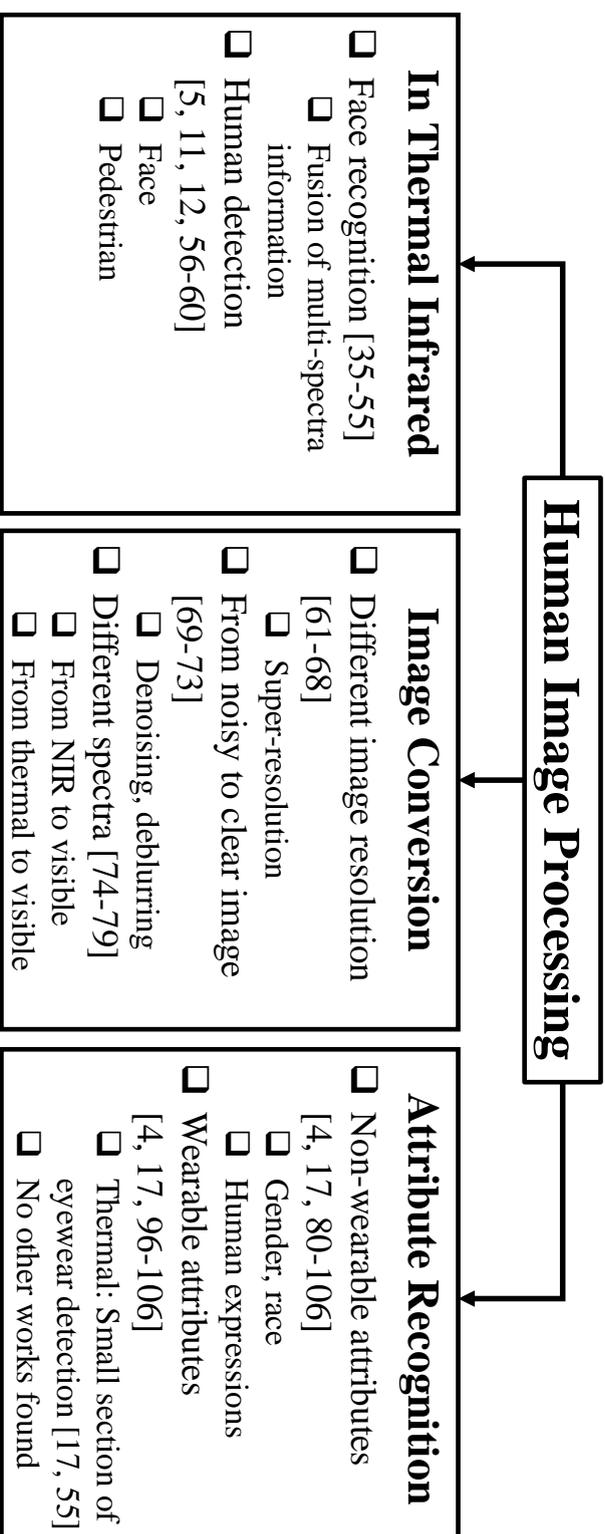


Figure 3.1: Overview of the related work.

## 3.1 Human Image Processing in the Thermal Infrared Spectrum

It has been established previously that thermal infrared cameras are capable of capturing human images regardless of the lighting/illumination conditions. This is the main advantage of using these types of cameras. However, it is difficult for humans to identify an individual based on their features in the thermal infrared spectrum. Consequently, many image processing researches conducted in this spectrum are related to researches for face recognition.

The earliest work on face recognition in the thermal infrared spectrum was done by Prokoski et al. [35] in 1992. There is a large amount of researches that has been done since, and the literature is still growing. In order to condense the literature, more attention is paid to some survey papers on this topic [36–40]. These papers describe various researches and their general approaches on face recognition in the thermal infrared spectrum. In their survey, Ghiass et al. [40] categorized the face recognition methodologies into four main groups: holistic appearance based, feature based, multi-spectral based, and multi-modal fusion based. The next related works are introduced according to this categorization.

The holistic appearance based method includes some earlier works [35] and work by Socolinsky et al. [41] which produced results with more than satisfactory recognition rates. Other researches which detect facial features such as eyes, nostrils, and mouths are also categorized in this group [42, 43]. In feature based methods, various features are first extracted before face recognition is performed. For example, feature extraction using Local Binary Patterns (LBP) [44] or wavelet transforms [45]. Unique to the thermal infrared, there are some feature based methods to extract vascular networks and blood perfusions based on the temperature [46–49]. Although there is no concrete evidence that the extracted features are in fact blood vessels, the representation of this “network” is special to the thermal infrared spectrum. An example of multi-spectral based face recognition methods was proposed by Bourlai

et al. [50], where they matched Mid-Wavelength InfraRed (MWIR) images to their visible images. Lastly, the multi-modal fusion methods have gained a lot of attraction recently. By fusing both visible and thermal infrared images, the advantages from both spectra can be combined. The fusion process is usually performed either at pixel-level or feature-level. This is a promising approach and many researches have been done, such as work by Arandjelovic et al. [51], Bhowmik et al. [52–54], and Kong et al. [55].

Besides face recognition, thermal infrared images are very useful for detection purposes because in most cases, human body temperature is higher than other objects in the environment. Many researches took form based on this observation such as work by Wong et al. [8, 9, 56]. In one of them, a face detection method was proposed by exploiting the geometric shape of human head [56]. Correa et al. [57] combined the skin information from the visible spectrum with thermal information from the thermal infrared spectrum to detect humans. Davis and Sharma [2, 58] detected humans with a background-subtraction approach, while others used the Histogram of Oriented Gradients (HOG) [59, 60].

## 3.2 Face Image Reconstruction

The subject of face image reconstruction is closely related with face image conversion, where many researches have been conducted. The conversion of face image can be performed from one image resolution to another in the same spectrum. This is a well-known research problem in the visible spectrum called the super-resolution problem, which is often referred to as face hallucination.

The term face hallucination was introduced by Baker and Kanade [61], where converting a very low-resolution face image to a high-resolution one seemed like a hallucination. Many researches followed suit in this topic, such as work by Liu et al. [62], Huang et al. [63], An and Bhanu [64], and Ma et al. [65] where they utilized the face image both globally and locally in small patches. Others used only the patches in

their methods [66, 67]. Additionally, a method to perform both the conversion and face recognition simultaneously has also been proposed [68].

Other than face hallucination, face deblurring and denoising can also be considered as image conversion problems. One of the work of face deblurring is by Pan et al. [69], where they proposed an exemplar-based algorithm to extract good structural information from exemplars. Nishiyama et al. [70] and Heflin et al. [71] inferred non-blurred facial appearances using blind image deconvolution. As for face denoising, works by Luo et al. [72] and Tang et al. [73] are among the many published researches on this subject.

Besides converting the image resolution, image conversion can also be performed from one spectrum to another. In comparison to the amount of publications available on face hallucination, this is not well explored. In common applications, the target of image conversion across the spectrum is the visible spectrum, providing face image that we humans can see. The spectrum which the image originated from can be any other spectrum, such as the subdivisions of the infrared spectrum.

There are researches aiming to reconstruct face images from the near-infrared spectrum to the visible spectrum. As mentioned previously, face images in the near-infrared spectrum resemble closely to those of the visible spectrum, especially in an ideal illumination condition. Therefore, the reconstruction process in this case is relatively simple. Shao et al. [74] approached the reconstruction problem by utilizing tensor spaces, while Chen et al. [75] and Zhang et al. [76] took a dictionary approach, collecting patches from face images in the training data. Goh et al. [77] utilized a model that studies the skin pigment based on the near infrared face images.

On the contrary, researches on reconstructing face images from the thermal infrared to the visible spectra are rare. To the extent of my knowledge, the only few works available on this topic are done by Li et al. [78] and Dou et al. [79]. Both of these works considered that local facial traits are important for the reconstruction, therefore they chose to employ a patch-based approach. The basis of their approach utilized Canonical Correlation Analysis (CCA) to learn the relationship between the

thermal infrared and the visible spectra. Li et al. [78] made use of Markov Random Fields (MRF) after the initial patch reconstruction, adjusting the patches for the global smoothness of the image. Dou et al. [79] on the other hand, proposed sophisticated Locally Linear Embedding (LLE), which tries to learn the changes of the different geometries between the thermal infrared and the visible spectra to reconstruct the patches. It can be concluded that the patch-based approach is considered as the state-of-the-art method in this subject. However, by only using the local facial traits, it has problems in regards to the reconstruction of the overall facial structure. This is because it does not consider the global structure of the face, which can make the reconstructed face look choppy and unnatural due to the disconnect between the patches.

### 3.3 Attribute Recognition

As mentioned previously in Chapter 1, attributes of a human encompass various aspects, from the ones directly correlated to their body (race, gender) to objects that they can wear (clothing, accessories). One of the earliest work on this subject was conducted by Golomb et al. [80] in 1990 where they introduced “SexNet”, a fully-connected back-propagation network trained to distinguish humans by their gender. Subsequently, there are many other works in recognizing non-wearable attributes. For example, recognition of human expressions by Bartlett et al. [81] and Lyons et al. [82], which both used Gabor filters for feature extraction. Other approaches have utilized different methods, such as a rich appearance dictionary [83] and Local Binary Patterns (LBP) [44] to achieve the recognition [84–86]. Mase [87] proposed the usage of optical flow to recognize facial expression from image sequences. For other purposes, LBP has also been used to recognize race/ethnicity [88]. Gutta et al. [89] and Shakhnarovich et al. [90] performed recognition on ethnicity and gender. Moghaddam and Yang [91] focused on gender recognition and utilized a Support Vector Machine (SVM) [92] to achieve it.

There are also some other interesting works related to attributes. For example, Ling et al. [93] performed a study on the effect of aging on face recognition. Sharma et al. [94] performed action recognition such as bike riding and running alongside with human attribute recognition. Christie et al. [95] attempted to predict user annoyance on image retrieval results based on its visual attributes.

In general, wearable attributes are commonly used in conjunction with its non-wearable counterpart for the attribute recognition research, for example, in the work by Bourdev et al. [96]. This can also be found in some works in attribute recognition from pedestrians [97]. Additionally, earlier works do not usually have attribute recognition as the end goal. For example, Kumar et al. [98, 99] utilized various attributes including the wearable attributes for face verification. They used various low-level features such as pixel intensity in RGB (Red, Green, and Blue) and HSV (Hue, Saturation, and Value) color spaces and edge magnitudes. Ijiri et al. [100] proposed a spatial codebook in their work on facial attribute recognition. Other works focused on searching people based on their attributes [1, 14, 101], or general attribute recognition for person re-identification [102–104]. Deep learning is also used for attribute recognition. For example, the “PANDA” network by Zhang et al. [105] performed attribute classification on images taken from Facebook, among others. Kang et al. [106] also used deep learning for facial attributes recognition in their work.

However, most of the aforementioned researches are conducted in the visible spectrum. In the thermal infrared spectrum, attribute recognition researches are very limited. For non-wearable attributes, Trujillo et al. [107] and Hernandez et al. [108] conducted researches on recognizing facial expressions in the thermal infrared spectrum.

In the subject of wearable attributes, however, there is only a small section of eyewear (glasses and sunglasses) detection that has been previously studied. Kong et al. [55] utilized ellipse fitting to detect glasses as part of preprocessing. Vaquero et al. [14] utilized Viola-Jones detector [109, 110] to perform eyewear detection in the infrared

spectrum, because of its unsatisfactory results in the visible spectrum. To the extent of my knowledge, there is no other work that has been conducted specifically on wearable attribute recognition in the thermal infrared domain.

It can be then concluded that the subject of wearable attribute recognition specifically is not widely explored. Additionally, the decomposition approach to extract the attributes proposed in this thesis is also novel.

## Chapter 4

# Visual Understanding from Thermal Infrared Image

There are two research topics in this thesis, which are each related to the understanding of human images in the thermal infrared spectrum. This chapter presents the first topic; Visual understanding from thermal infrared image. This chapter is structured as follows. Section 4.1 describes the general background of the face image reconstruction research and its relation to the theme of the thesis; Thermal infrared image understanding. Section 4.2 elaborates the proposed framework made to gain visual understanding of the human image in the thermal infrared spectrum. The approach to the problem is by investigating the correlation between the face images in both spectra. After establishing the correlation, it is possible to perform cross-modal face image reconstruction from thermal infrared to visible spectra. Next, the experiments along with its results and analysis are described in Section 4.3. Additional experiments for further discussion are also provided in Section 4.4 before the chapter is closed with the summary in Section 4.5.

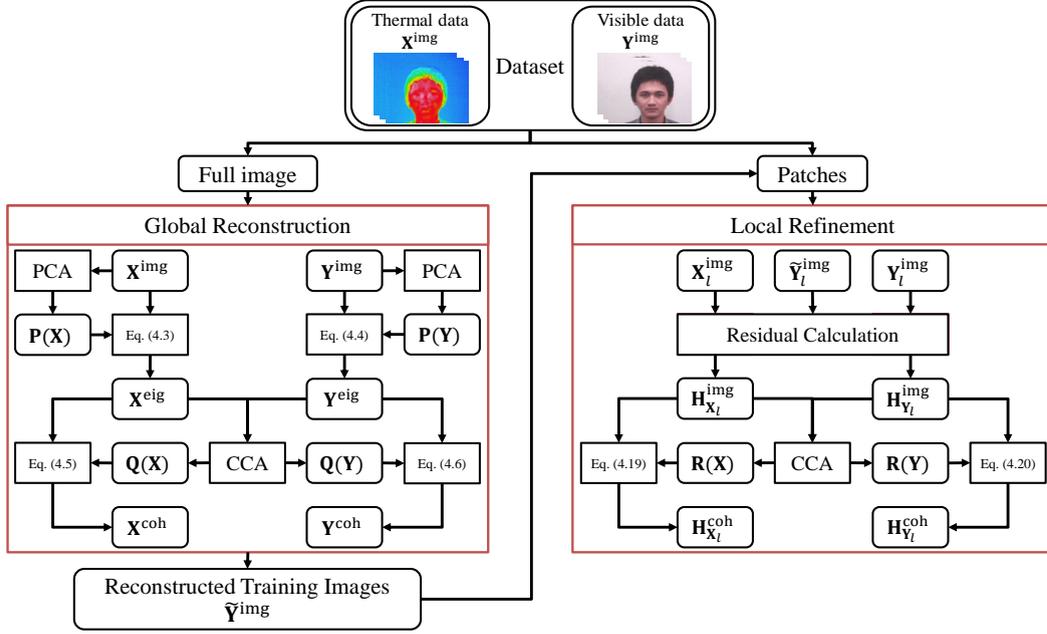
## 4.1 Background

In Chapter 1, the first research question in this thesis was discussed. The question was “What does the person look like in the visible spectrum?” The research topic corresponding to it proposes the reconstruction of face images from the thermal infrared spectrum to the visible spectrum to answer the question. The overview of the approach was explained in Section 1.3.1. Face image reconstruction provides visualization to humans, which is very useful for surveillance systems that employ thermal infrared cameras. Additionally, the results of the reconstruction can be used for further image processing such as face recognition.

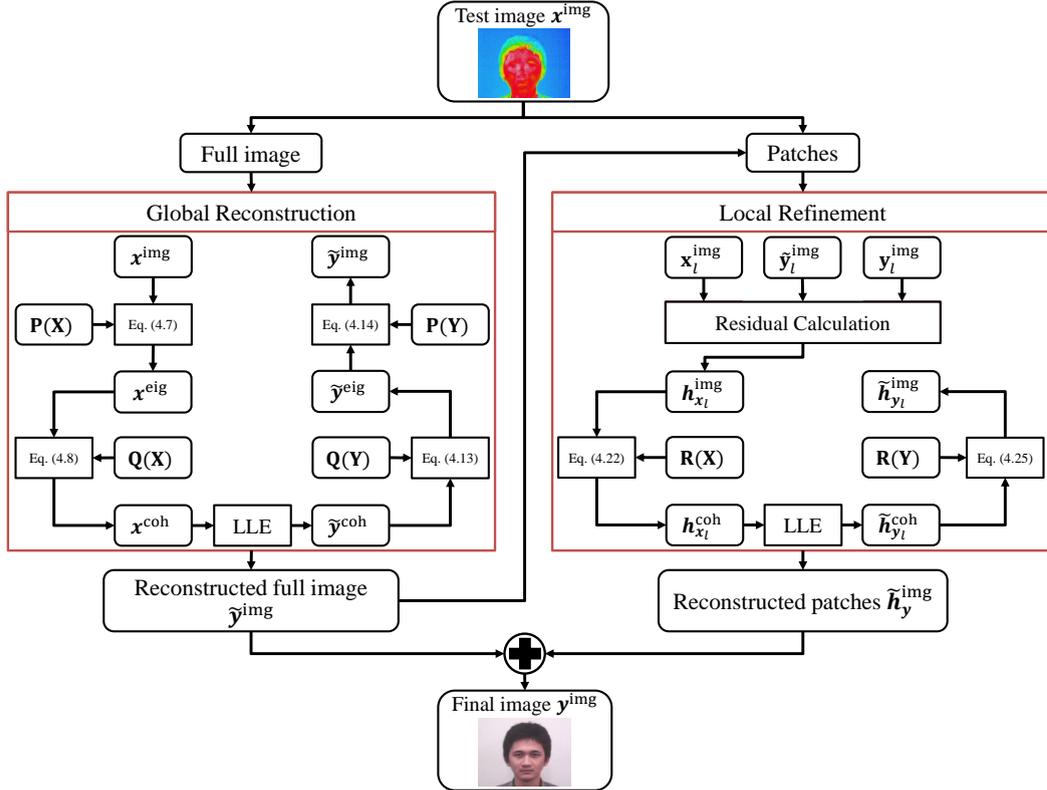
A system needs a certain understanding in order to be able to achieve the reconstruction. In this thesis, this understanding is referred to as the visual understanding. The visual understanding required is the correlation between thermal infrared and visible spectra human images. The research in this chapter uses this train of thought to propose a reconstruction framework.

## 4.2 Learning-based Reconstruction Framework

In this research, the proposed method employs a two-step reconstruction approach. The usage of this two-step method is novel in the reconstruction subject across different spectra, especially from the thermal infrared spectrum to the visible spectrum. The first step of this method is called the Global Reconstruction and the second step is called the Local Refinement. Each step involves its own training and reconstruction phases. The overall flow of the framework is shown in Figure 4.1. In order to provide a clear description of the framework, the explanation of the reconstruction method is divided based on these two steps.



(A) Training Phase



(B) Reconstruction Phase

FIGURE 4.1: Process flow of the proposed reconstruction method.

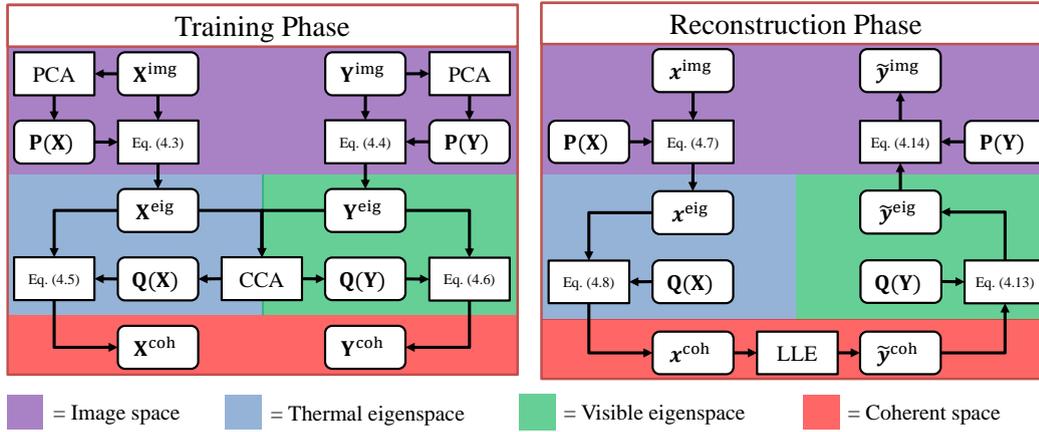


FIGURE 4.2: Overview of the Global Reconstruction step.

## 4.2.1 Global Reconstruction Step

In the Global Reconstruction step, a face image is reconstructed as a whole from the thermal infrared to the visible spectra. There are two phases in this step: The Training Phase prepares the system by establishing the relationship between the thermal infrared and the visible images using Canonical Correlation Analysis (CCA). The Reconstruction Phase utilizes Locally Linear Embedding (LLE) to reconstruct the face image based on the relationship obtained in the training phase. The overview of this step can be seen in Figure 4.2, which is equivalent to the left side of Figures 4.1(A) and 4.1(B).

### 4.2.1.1 Training Phase

Let's assume a dataset of known pairs of images in the thermal infrared and visible spectra, with thermal infrared images represented by  $\mathbf{X}^{\text{space}} = [\mathbf{x}_1^{\text{space}} \ \mathbf{x}_2^{\text{space}} \ \dots \ \mathbf{x}_N^{\text{space}}]$ , and the visible images represented by  $\mathbf{Y}^{\text{space}} = [\mathbf{y}_1^{\text{space}} \ \mathbf{y}_2^{\text{space}} \ \dots \ \mathbf{y}_N^{\text{space}}]$ . The superscript "space" indicates the space where the data is located, such as image space "img," eigenspace "eig," and coherent space "coh."  $N$  represents the total number of training data.

First, both the thermal infrared and the visible face images are centered according to the following:

$$\widehat{\mathbf{x}}_n^{\text{img}} = \mathbf{x}_n^{\text{img}} - \boldsymbol{\mu}_{\mathbf{X}}, \quad (4.1)$$

$$\widehat{\mathbf{y}}_n^{\text{img}} = \mathbf{y}_n^{\text{img}} - \boldsymbol{\mu}_{\mathbf{Y}}. \quad (4.2)$$

Note that the centering is performed on each data point  $n = 1, 2, \dots, N$ , resulting in  $\widehat{\mathbf{X}}^{\text{img}} = [\widehat{\mathbf{x}}_1^{\text{img}} \widehat{\mathbf{x}}_2^{\text{img}} \dots \widehat{\mathbf{x}}_N^{\text{img}}]$  and  $\widehat{\mathbf{Y}}^{\text{img}} = [\widehat{\mathbf{y}}_1^{\text{img}} \widehat{\mathbf{y}}_2^{\text{img}} \dots \widehat{\mathbf{y}}_N^{\text{img}}]$  for centered thermal and visible data, respectively. The notation  $\widehat{\cdot}$  indicates that the average has been subtracted from the data. The mean vectors  $\boldsymbol{\mu}_{\mathbf{X}}$  and  $\boldsymbol{\mu}_{\mathbf{Y}}$  of each spectrum will be used later.

The centered training data  $\widehat{\mathbf{X}}^{\text{img}}$  and  $\widehat{\mathbf{Y}}^{\text{img}}$  are then projected onto the eigenspace by utilizing projection matrices  $\mathbf{P}(\mathbf{X})$  and  $\mathbf{P}(\mathbf{Y})$  as shown in Equations (4.3) and (4.4). The projection matrices are obtained from the usage of Principal Component Analysis (PCA) on the training data. It is important to note that the thermal and the visible data are each projected onto their own eigenspaces  $\mathbf{X}^{\text{eig}} = [\mathbf{x}_1^{\text{eig}} \mathbf{x}_2^{\text{eig}} \dots \mathbf{x}_N^{\text{eig}}]$  and  $\mathbf{Y}^{\text{eig}} = [\mathbf{y}_1^{\text{eig}} \mathbf{y}_2^{\text{eig}} \dots \mathbf{y}_N^{\text{eig}}]$  for the thermal and visible data, respectively.

$$\mathbf{X}^{\text{eig}} = (\mathbf{P}(\mathbf{X}))^T \widehat{\mathbf{X}}^{\text{img}} \quad (4.3)$$

$$\mathbf{Y}^{\text{eig}} = (\mathbf{P}(\mathbf{Y}))^T \widehat{\mathbf{Y}}^{\text{img}} \quad (4.4)$$

With the thermal and visible training data in their separate eigenspaces, the relationship between them can be studied by applying CCA. In short, CCA is used to find an index describing the link between the two datasets [111].

First, the proposed method uses CCA to produce two projection matrices, one for each type of data. These projection matrices,  $\mathbf{Q}(\mathbf{X})$  and  $\mathbf{Q}(\mathbf{Y})$ , are used to project the thermal and visible training data from their own eigenspaces onto a coherent space, as shown in the following:

$$\mathbf{X}^{\text{coh}} = (\mathbf{Q}(\mathbf{X}))^T \mathbf{X}^{\text{eig}}, \quad (4.5)$$

$$\mathbf{Y}^{\text{coh}} = (\mathbf{Q}(\mathbf{Y}))^T \mathbf{Y}^{\text{eig}}, \quad (4.6)$$

where  $\mathbf{X}^{\text{coh}}$  and  $\mathbf{Y}^{\text{coh}}$  represent the training data of both thermal and visible spectrum in the coherent space, respectively. At this point, the proposed method has projection matrices from the image space onto two different eigenspaces  $\mathbf{P}(\mathbf{X})$  and  $\mathbf{P}(\mathbf{Y})$ , and from the two eigenspaces onto a coherent space  $\mathbf{Q}(\mathbf{X})$  and  $\mathbf{Q}(\mathbf{Y})$ . Also, it keeps the mean vectors  $\boldsymbol{\mu}_X$  and  $\boldsymbol{\mu}_Y$  for further use in the reconstruction phase. This marks the end of the training phase of the Global Reconstruction step.

#### 4.2.1.2 Reconstruction Phase

In this phase of the Global Reconstruction step, the proposed method reconstructs a new thermal face image  $\mathbf{x}^{\text{img}}$  to a visible face image  $\tilde{\mathbf{y}}^{\text{img}}$ . The notation  $\tilde{\cdot}$  indicates reconstructed data. To achieve this, first, it needs to project the new thermal face image onto the coherent space. There are two operations performed for this purpose, starting with the projection of the thermal face image  $\mathbf{x}^{\text{img}}$  onto the eigenspace as shown as follows:

$$\mathbf{x}^{\text{eig}} = (\mathbf{P}(\mathbf{X}))^T (\mathbf{x}^{\text{img}} - \boldsymbol{\mu}_X), \quad (4.7)$$

where  $\mathbf{P}(\mathbf{X})$  is the projection matrix from image space to thermal infrared's eigenspace and  $\boldsymbol{\mu}_X$  is the mean vector for thermal data in the image space. The second operation is the projection of the face data  $\mathbf{x}^{\text{eig}}$  to the coherent space shown in Equation (4.8), where  $\mathbf{x}^{\text{coh}}$  is the projected thermal face image in the coherent space,  $\mathbf{Q}(\mathbf{X})$  is the projection matrix from thermal infrared's eigenspace to coherent space. In Equations (4.7) and (4.8),  $\boldsymbol{\mu}_X$ ,  $\mathbf{P}(\mathbf{X})$ , and  $\mathbf{Q}(\mathbf{X})$  are obtained from the training phase.

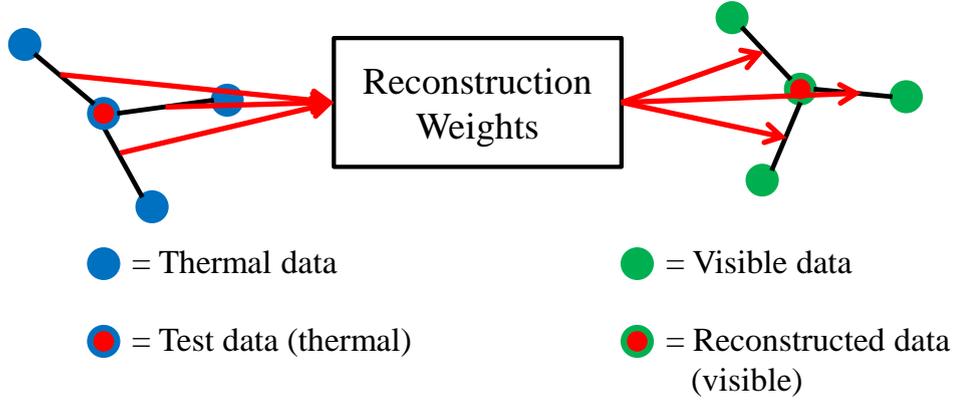


FIGURE 4.3: Overview of the LLE method.

$$\mathbf{x}^{\text{coh}} = (\mathbf{Q}(\mathbf{X}))^T \mathbf{x}^{\text{eig}} \quad (4.8)$$

The reconstruction phase is performed in the coherent space by Locally Linear Embedding (LLE) [17], a neighbor-based reconstruction method which requires a certain number of neighbors to do so.

A simple overview of LLE can be seen in Figure 4.3. The LLE starts by using the nearest neighbor method to find  $K$  neighbors of  $\mathbf{x}^{\text{coh}}$ , represented by  $\mathbf{A}_x = [\mathbf{a}_x^1 \ \mathbf{a}_x^2 \ \cdots \ \mathbf{a}_x^K]$ . Further process involves the minimization of the error function as follows:

$$\epsilon(\mathbf{w}) = \left| \mathbf{x}^{\text{coh}} - \sum_{k=1}^K w^k \mathbf{a}_x^k \right|, \quad (4.9)$$

where  $\mathbf{w} = (w^1, w^2, \dots, w^K)^T$  is the reconstruction weight vector. The minimization problem can be solved by introducing a local gram matrix  $\mathbf{G}$  with  $\mathbf{G}_{j,k}$  as its element where  $j, k = 1, 2, \dots, K$  as shown in Equation (4.10).

$$\mathbf{G}_{j,k} = (\mathbf{x}^{\text{coh}} - \mathbf{a}_x^j) \cdot (\mathbf{x}^{\text{coh}} - \mathbf{a}_x^k) \quad (4.10)$$

The weight vector  $\mathbf{w}$  can be calculated with Equation (4.11), where  $\mathbf{G}_{j,k}^{-1}$  is an element of  $\mathbf{G}^{-1}$  (the inverse of matrix  $\mathbf{G}$ ). The reconstruction error is minimized by the use of the Lagrange multiplier to enforce  $\sum_k w^k = 1$  [112].

$$w^k = \frac{\sum_j \mathbf{G}_{j,k}^{-1}}{\sum_j \sum_k \mathbf{G}_{j,k}^{-1}} \quad (4.11)$$

Utilizing the weight vector  $\mathbf{w}$ , the visible version of the data can be estimated as follows:

$$\tilde{\mathbf{y}}^{\text{coh}} = \sum_{k=1}^K w^k \mathbf{a}_y^k, \quad (4.12)$$

where  $\mathbf{A}_y = [\mathbf{a}_y^1 \ \mathbf{a}_y^2 \ \cdots \ \mathbf{a}_y^K]$  has a configuration identical to that of  $\mathbf{A}_x$ . This means that when  $\mathbf{x} = \mathbf{y}$ ,  $\mathbf{A}_x$  and  $\mathbf{A}_y$  refer to thermal infrared and visible data of one pair, respectively.

After the new visible data  $\tilde{\mathbf{y}}^{\text{coh}}$  has been reconstructed in the coherent space, the proposed method can then project the data back onto the image space. Similar to how it projects an image from the image space onto the coherent space, there are two operations to perform the reverse operation. The proposed method projects the data from the coherent space onto the eigenspace of visible spectrum based on the following equation:

$$\tilde{\mathbf{y}}^{\text{eig}} = (\mathbf{Q}(\mathbf{Y}))^\dagger \tilde{\mathbf{y}}^{\text{coh}}, \quad (4.13)$$

where  $(\mathbf{Q}(\mathbf{Y}))^\dagger = (\mathbf{Q}(\mathbf{Y})(\mathbf{Q}(\mathbf{Y}))^T)^{-1}\mathbf{Q}(\mathbf{Y})$  is the pseudo-inverse of the projection matrix. The next process is projecting the reconstructed data from the visible spectrum's eigenspace onto the image space as follows:

$$\tilde{\mathbf{y}}^{\text{img}} = \mathbf{P}(\mathbf{Y})\tilde{\mathbf{y}}^{\text{eig}} + \mu_Y \quad (4.14)$$

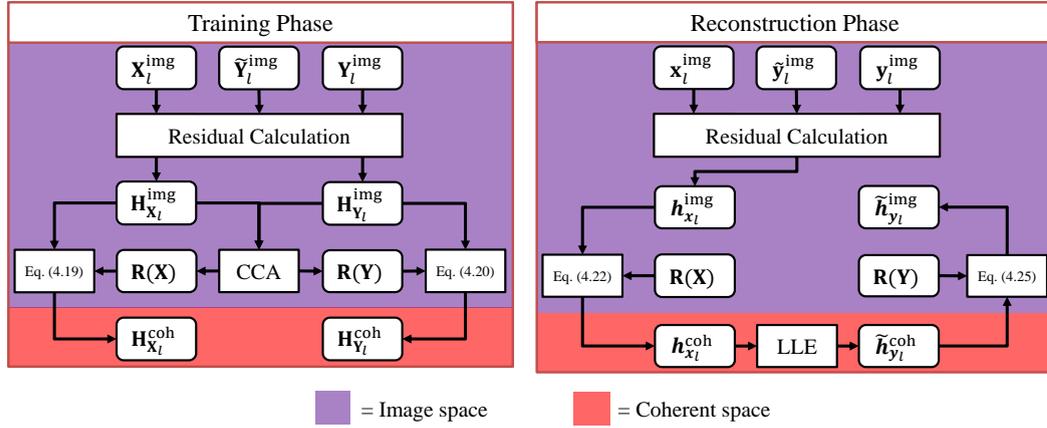


FIGURE 4.4: Overview of the Local Refinement step.

where  $\mu_Y$  is the mean vector for visible data in the visible spectrum and  $\mathbf{P}(Y)$  is the projection matrix. Note that  $\mu_Y$  is the mean vector from Equation (4.2). With  $\tilde{\mathbf{y}}^{\text{img}}$  representing the globally reconstructed visible face image, the reconstruction phase of the Global Reconstruction step is concluded.

## 4.2.2 Local Refinement Step

As mentioned previously, this step involves multiple reconstructions of patches of the face image. The core idea of the Local Refinement step is the usage of the residual component to refine the reconstructed data from the Global Reconstruction step. The overview of this step can be seen in Figure 4.4, which is equivalent to the right side of Figures 4.1(A) and 4.1(B). The patches involved are obtained by a sliding window with overlapping pixel information. Visual examples of the patches can be seen in Figure 4.5.

### 4.2.2.1 Training Phase

In order to proceed with this step, the globally reconstructed images of training data are needed. The notation for the globally reconstructed training data is  $\tilde{\mathbf{Y}}_{n,\ell}^{\text{img}} = [\tilde{\mathbf{y}}_{1,\ell}^{\text{img}} \tilde{\mathbf{y}}_{2,\ell}^{\text{img}} \cdots \tilde{\mathbf{y}}_{N,\ell}^{\text{img}}]$ , where  $n = 1, 2, \dots, N$  represents the index of each observation

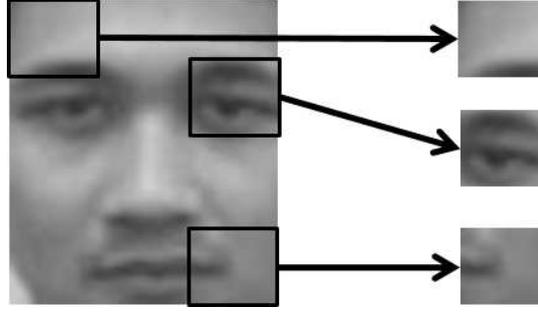


FIGURE 4.5: Examples of patches taken from a face image.

and  $\ell = 1, 2, \dots, L$  represents the index of each patch. First, the residual component between the reconstructed data and the actual training data  $\mathbf{Y}_{n,\ell}^{\text{img}} = [\mathbf{y}_{1,\ell}^{\text{img}} \ \mathbf{y}_{2,\ell}^{\text{img}} \ \dots \ \mathbf{y}_{N,\ell}^{\text{img}}]$  is calculated as follows:

$$\mathbf{h}_{\mathbf{y}_{n,\ell}}^{\text{img}} = \mathbf{y}_{n,\ell}^{\text{img}} - \tilde{\mathbf{y}}_{n,\ell}, \quad (4.15)$$

where  $\mathbf{H}_{\mathbf{Y}_\ell}^{\text{img}} = [\mathbf{h}_{\mathbf{y}_{1,\ell}}^{\text{img}} \ \mathbf{h}_{\mathbf{y}_{2,\ell}}^{\text{img}} \ \dots \ \mathbf{h}_{\mathbf{y}_{N,\ell}}^{\text{img}}]$  represents the residual components. As for the thermal data, the residual component is calculated as follows:

$$\mathbf{h}_{\mathbf{x}_{n,\ell}}^{\text{img}} = \mathbf{x}_{n,\ell}^{\text{img}} - \tilde{\mathbf{y}}_{n,\ell}, \quad (4.16)$$

where the residual component of the overall thermal data is represented by  $\mathbf{H}_{\mathbf{X}_\ell}^{\text{img}} = [\mathbf{h}_{\mathbf{x}_{1,\ell}}^{\text{img}} \ \mathbf{h}_{\mathbf{x}_{2,\ell}}^{\text{img}} \ \dots \ \mathbf{h}_{\mathbf{x}_{N,\ell}}^{\text{img}}]$ .

It is important to note that the residual components from Equation (4.16) contains information from the globally reconstructed image instead of information only from the original thermal image, therefore enriching the information available for the training. Before moving on to the next process, the training patches are centered according to Equations (4.17) and (4.18) for thermal infrared and visible data, respectively. Note that  $\lambda_{\mathbf{X}_\ell}$  and  $\lambda_{\mathbf{Y}_\ell}$  represent the average of patch  $\ell$  in the thermal and the visible data, respectively.

$$\widehat{\mathbf{h}}_{\mathbf{x}_{n,\ell}}^{\text{img}} = \mathbf{h}_{\mathbf{x}_{n,\ell}}^{\text{img}} - \lambda_{\mathbf{X}_\ell} \quad (4.17)$$

$$\widehat{\mathbf{h}}_{\mathbf{y}_{n,\ell}}^{\text{img}} = \mathbf{h}_{\mathbf{y}_{n,\ell}}^{\text{img}} - \boldsymbol{\lambda}_{\mathbf{Y}_\ell} \quad (4.18)$$

The results of these centering operations are  $\widehat{\mathbf{H}}_{\mathbf{X}_\ell}^{\text{img}} = [\widehat{\mathbf{h}}_{x_{1,\ell}}^{\text{img}} \widehat{\mathbf{h}}_{x_{2,\ell}}^{\text{img}} \cdots \widehat{\mathbf{h}}_{x_{N,\ell}}^{\text{img}}]$  for thermal data and  $\widehat{\mathbf{H}}_{\mathbf{Y}_\ell}^{\text{img}} = [\widehat{\mathbf{h}}_{y_{1,\ell}}^{\text{img}} \widehat{\mathbf{h}}_{y_{2,\ell}}^{\text{img}} \cdots \widehat{\mathbf{h}}_{y_{N,\ell}}^{\text{img}}]$  for visible data. The centered thermal data  $\widehat{\mathbf{H}}_{\mathbf{X}_\ell}^{\text{img}}$  and centered visible data  $\widehat{\mathbf{H}}_{\mathbf{Y}_\ell}^{\text{img}}$  are then projected onto the coherent space as follows:

$$\mathbf{H}_{\mathbf{X}_\ell}^{\text{coh}} = (\mathbf{R}(\mathbf{X}_\ell))^T \widehat{\mathbf{H}}_{\mathbf{X}_\ell}^{\text{img}} \quad (4.19)$$

$$\mathbf{H}_{\mathbf{Y}_\ell}^{\text{coh}} = (\mathbf{R}(\mathbf{Y}_\ell))^T \widehat{\mathbf{H}}_{\mathbf{Y}_\ell}^{\text{img}} \quad (4.20)$$

The projection matrices  $\mathbf{R}(\mathbf{X}_\ell)$  and  $\mathbf{R}(\mathbf{Y}_\ell)$  are obtained from CCA, and the training patches in the coherent space are represented by  $\mathbf{H}_{\mathbf{X}_\ell}^{\text{coh}}$  and  $\mathbf{H}_{\mathbf{Y}_\ell}^{\text{coh}}$ . Similar to the Global Reconstruction step, the training phase concludes here, when the training data of both thermal infrared and visible spectra are represented in the coherent space.

#### 4.2.2.2 Reconstruction Phase

The processes of reconstruction in this step involve multiple small patches of an image. The first operation in this phase is to calculate the residual component  $\mathbf{h}_x^{\text{img}}$  from the globally reconstructed data  $\widehat{\mathbf{y}}^{\text{img}}$  as follows:

$$\mathbf{h}_x^{\text{img}} = \mathbf{x}^{\text{img}} - \widehat{\mathbf{y}}^{\text{img}}. \quad (4.21)$$

For the next operation, it is necessary to retrieve patches from the residual component  $\mathbf{h}_x^{\text{img}}$ , which are represented as  $\mathbf{h}_{x_\ell}^{\text{img}}$  where  $\ell = 1, 2, \dots, L$  is the patch index. These patches are projected onto the coherent space as follows:

$$\mathbf{h}_{x_\ell}^{\text{coh}} = (\mathbf{R}(\mathbf{X}_\ell))^T (\mathbf{h}_{x_\ell}^{\text{img}} - \boldsymbol{\lambda}_{\mathbf{X}_\ell}), \quad (4.22)$$

where projection matrix  $\mathbf{R}(\mathbf{X}_\ell)$  and mean vector  $\lambda_{\mathbf{X}_\ell}$  are obtained from the training phase. With the residual patches in the coherent space, the proposed method can reconstruct new residual patches by LLE. In this case, the minimization problem that needs to be solved is the following:

$$\epsilon(\mathbf{w}) = |\mathbf{h}_{\mathbf{x}_\ell}^{\text{coh}} - \sum_{k=1}^K w_\ell^k \mathbf{b}_{\mathbf{x}_\ell}^k|, \quad (4.23)$$

where  $\mathbf{w} = (w_\ell^1, w_\ell^2, \dots, w_\ell^K)^T$  is the weight vector for the reconstruction and  $\mathbf{B}_{\mathbf{x}_\ell} = [\mathbf{b}_{\mathbf{x}_\ell}^1 \mathbf{b}_{\mathbf{x}_\ell}^2 \dots \mathbf{b}_{\mathbf{x}_\ell}^K]$  represents the  $K$  nearest neighbors of  $\mathbf{h}_{\mathbf{x}_\ell}^{\text{coh}}$  at patch  $\ell$ . The reconstruction weight vector can be calculated the same way they were obtained in the Global Reconstruction step. The reconstructed residual patch for the visible spectrum is calculated after the proposed method have obtained the reconstruction weight vector as follows:

$$\tilde{\mathbf{h}}_{\mathbf{y}_\ell}^{\text{coh}} = \sum_{k=1}^K w_\ell^k \mathbf{b}_{\mathbf{y}_\ell}^k, \quad (4.24)$$

where  $\mathbf{B}_{\mathbf{y}_\ell} = [\mathbf{b}_{\mathbf{y}_\ell}^1 \mathbf{b}_{\mathbf{y}_\ell}^2 \dots \mathbf{b}_{\mathbf{y}_\ell}^K]$  represents  $K$  neighbors whose indices are the same with that of  $\mathbf{B}_{\mathbf{x}_\ell}$ . It is important to note that the reconstruction weight vector is different for each patch location.

The next operation is to project these reconstructed patches back onto the image space as follows:

$$\tilde{\mathbf{h}}_{\mathbf{y}_\ell}^{\text{img}} = (\mathbf{R}(\mathbf{Y}_\ell))^\dagger \tilde{\mathbf{h}}_{\mathbf{y}_\ell}^{\text{coh}} + \lambda_{\mathbf{Y}_\ell}, \quad (4.25)$$

where  $(\mathbf{R}(\mathbf{Y}_\ell))^\dagger = (\mathbf{R}(\mathbf{Y})(\mathbf{R}(\mathbf{Y}))^T)^{-1} \mathbf{R}(\mathbf{Y})$  is the pseudo-inverse of the projection matrix. Before the final image is created, the reconstructed patches are combined into one residual image  $\tilde{\mathbf{h}}_{\mathbf{y}}^{\text{img}}$  by averaging the overlapping pixels of the patches  $\tilde{\mathbf{h}}_{\mathbf{y}_\ell}^{\text{img}}$ . The final image is created by combining the residual image and the globally reconstructed image as follows:

$$\mathbf{y}^{\text{img}} = \widetilde{\mathbf{y}}^{\text{img}} + \widetilde{\mathbf{h}}_y^{\text{img}}. \quad (4.26)$$

For simplicity, the final image  $\mathbf{y}^{\text{img}}$  does not have the  $\widetilde{\cdot}$  notation. This concludes the reconstruction phase of the Local Refinement and the reconstruction framework as a whole.

## 4.3 Experiments and Analysis

To assess the reconstruction capability of the proposed method, experiments were conducted by evaluating the produced face images. This section starts by introducing the dataset used for the experiment, followed by the setup of the experiment elaborated in the following subsection. The next subsection describes some other methods as comparison. Finally, the reconstruction results and the analysis are provided in the last subsection.

### 4.3.1 Dataset

Since a public thermal infrared image dataset that contains a high number of people in both thermal infrared and visible spectra was not readily available, a private dataset was newly created for this research. It consists of face images of 180 Japanese people (169 males and 11 females in the age range of 19–54 years old) with five subtle variations for each person, taken simultaneously in both the thermal infrared and the visible spectra. Subjects were standing at a distance of 1.0 to 1.5 m from the camera, and only the head region of the body was captured by the camera.

A total of 1,800 images were captured, with 900 images in each spectrum. These images were taken indoors at room temperature (around 22–25 degrees Celsius), using an Avionics' TVS-500EX camera [113], which can capture infrared wavelengths

from 8 to 14  $\mu\text{m}$ . The camera was set to capture infrared radiation using its automatic function, which means it adaptively sets the minimum and the maximum temperature according to the scene. The captured images show that the minimum temperature ranges from 32 to 34 degrees Celsius while the maximum temperature ranges from 36 to 38 degrees Celsius. The camera is able to take pictures from both the thermal infrared and the visible spectra simultaneously, although the focal axes are slightly different. The raw image size was  $320 \times 240$  pixels for the thermal infrared spectrum and  $640 \times 480$  pixels for the visible spectrum.

Various preprocessings were performed to these pairs of images. First, the eyes and the mouth of each person were manually annotated. Given these annotations, affine transformation was performed on the thermal infrared images to align themselves with their visible pairs. Then, the size of these face images (both thermal and visible) were down-sampled to  $56 \times 64$  pixels.

### 4.3.2 Experimental Setup

In this experiment, it is assumed that for every thermal face test image, variations of their face image exist in the training data. This means that from multiple face variations of a person, one is used for testing while the rest of the variations are used for training. Therefore, the images used for training and testing do not intersect. The experiment was performed with cross-validation.

To evaluate the reconstructed face images, Peak Signal to Noise Ratio (PSNR) [114] and Structural Similarity (SSIM) [115] were used as measures. Both of these measures compare the reconstructed face with its original image in the visible spectrum. These measures are considered as means to evaluate the quality of the reconstructed face.

Let us assume a ground truth image  $A$  and a reconstructed image  $B$  where the size of the image is  $m \times n$  [pixels]. The PSNR value is calculated as follows:

$$\text{PSNR} = 20 \times \log_{10} 255 - 10 \times \log_{10} \text{MSE}, \quad (4.27)$$

where MSE stands for Mean Squared Error, defined as follows:

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (A(i, j) - B(i, j))^2. \quad (4.28)$$

The SSIM value is based on three terms, namely luminance, contrast, and structure. SSIM is calculated by a weighted combination of these terms as follows:

$$\text{SSIM} = (l(A, B))^\alpha (c(A, B))^\beta (s(A, B))^\gamma, \quad (4.29)$$

where default values of  $\alpha$ ,  $\beta$ , and  $\gamma$  are 1. Each term is calculated as follows:

$$\begin{aligned} l(A, B) &= \frac{2\mu_A\mu_B + c_1}{\mu_A^2 + \mu_B^2 + c_1}, \\ c(A, B) &= \frac{2\sigma_A\sigma_B + c_2}{\sigma_A^2 + \sigma_B^2 + c_2}, \\ s(A, B) &= \frac{2\sigma_{AB} + c_3}{\sigma_A\sigma_B + c_3}, \end{aligned} \quad (4.30)$$

where  $\mu_A$  and  $\mu_B$  represent averages,  $\sigma_A$  and  $\sigma_B$  represent standard deviations, and  $\sigma_{AB}$  represents covariance.

In addition to PSNR and SSIM, face recognition was conducted using the EigenFace method [116], to further evaluate the usability of the reconstructed face images.

The proposed framework involves several parameters that need to be considered in order to perform reconstruction of a face image, such as the number of neighbors used by the LLE method or the size of the patch in the Local Refinement step retrieved from a face image. For this particular experiment, the proposed method used

TABLE 4.1: Comparison of the methods performed in the experiment.

Method	Whole Image	Patch	Reconstruction method
Proposed	✓	✓	LLE
Holistic LLE	✓	—	LLE
Patch-Based LLE	—	✓	LLE
Patch-Based 1-NN	—	✓	NN
Patch-Based $k$ -NN	—	✓	NN

five nearest neighbors for LLE, and a patch size of  $9 \times 9$  pixels. The parameters were decided empirically through an experiment described in Section 4.4.1.

### 4.3.3 Comparative Methods

The experiment compares the capability of the proposed two-step method with holistic only and patch-based only methods. The experiment was conducted in the same way for all of the methods to guarantee fairness. When applicable, the parameters used were also the same. The comparative methods used in this experiment are provided in the following list. Table 4.1 shows the comparison of the methods performed in this experiment.

- **Holistic LLE:** Reconstruction is performed utilizing only the whole image, without considering the local traits. In other words, this method only applies the Global Reconstruction step without the Local Refinement step.
- **Patch-Based LLE:** Reconstruction is performed by applying the patch-based method directly on the image instead of the residual image. In other words, this method only applies the Local Refinement step without the Global Reconstruction step.
- **Patch-Based 1-NN:** Reconstruction is performed by first finding the most similar thermal patch from the training data. Then, the face is reconstructed using

TABLE 4.2: Comparison of the proposed method with the holistic method and various patch-based methods (Number of neighbors: 5, Patch size:  $9 \times 9$  pixels). SE stands for the standard error, and is represented in the bracket.

Method	PSNR (SE) [dB]	SSIM (SE)	Recog. Rate [%]
Proposed	<b>33.11</b> (3.69)	<b>0.95</b> (0.05)	<b>98.44</b>
Holistic LLE	27.04 (1.81)	0.85 (0.05)	98.33
Patch-Based LLE	29.21 (4.09)	0.92 (0.06)	87.33
Patch-Based 1-NN	19.47 (1.78)	0.73 (0.06)	1.45
Patch-Based $k$ -NN	25.38 (3.41)	0.88 (0.07)	63.78

the visible pair of the said patch. In other words, this is the most conventional reconstruction method. The NN in Patch-Based 1-NN stands for Nearest Neighbor(s).

- **Patch-Based  $k$ -NN:** Reconstruction is performed by first finding the  $k$  most similar thermal patches from the training data. Then, the face is reconstructed using the average of the visible pair of said patches. In other words, it considers  $k$  visible patches for the reconstruction, where  $k$  is the number of nearest neighbors.

#### 4.3.4 Results and Analysis

The evaluations of the proposed method, the holistic method, and the various patch-based methods are summarized in Table 4.2. Bar graph representations are also shown in Figures 4.6, 4.7, and 4.8 for the PSNR, SSIM, and recognition rate, respectively. Actual reconstruction results of these methods can be seen in Figure 4.9. A heat map representation of the confusion matrix of the recognition evaluation is also provided, specifically for the results of the proposed method in Figure 4.10. Visual examples of the reconstruction by the proposed method throughout its process are also provided in Figure 4.11.

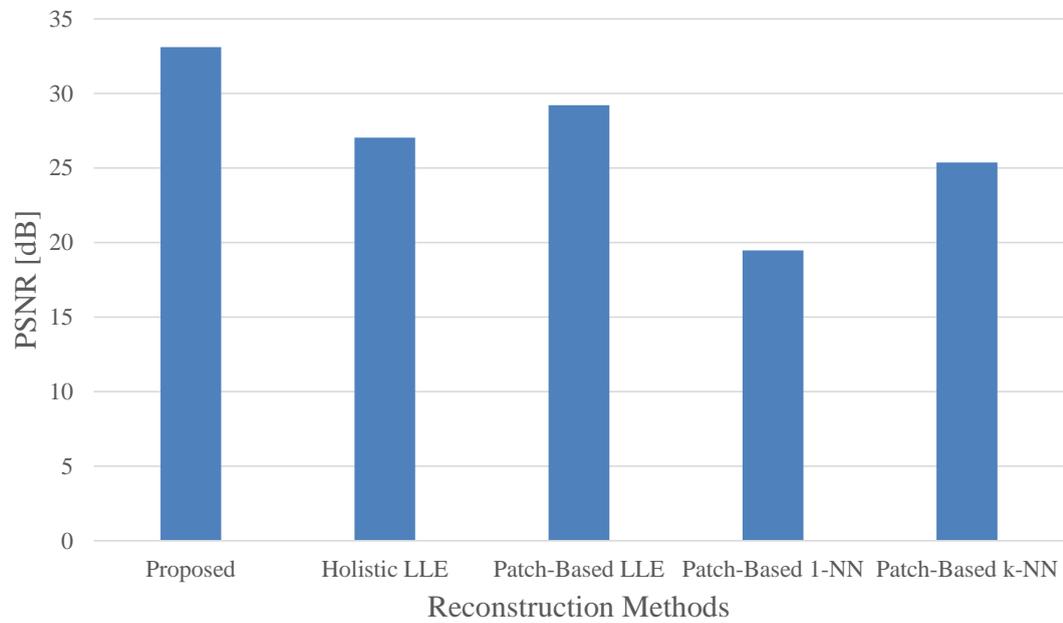


FIGURE 4.6: Evaluation results of PSNR.

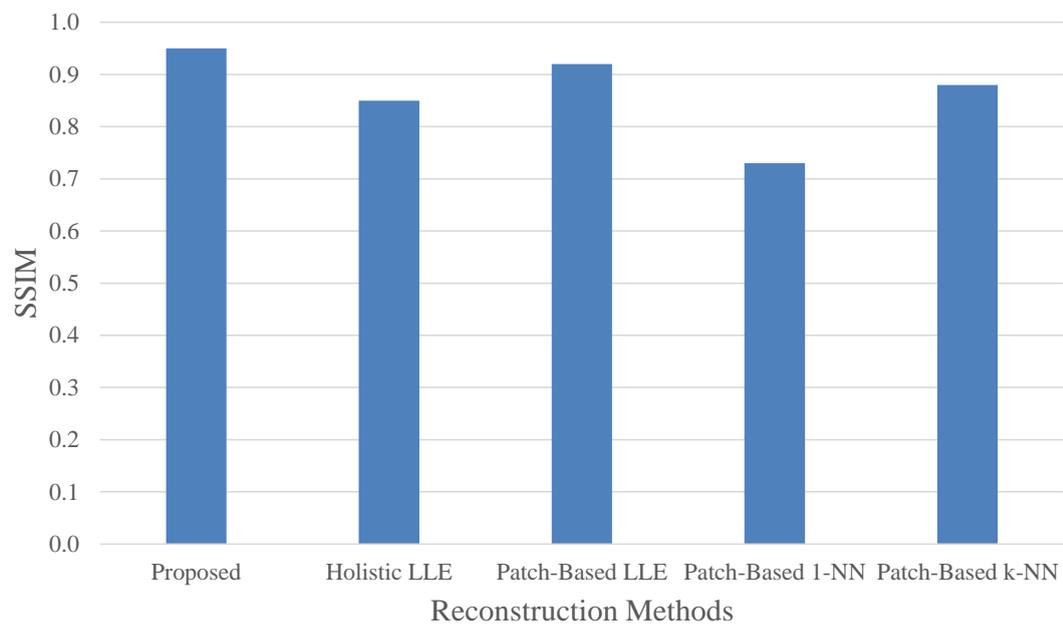


FIGURE 4.7: Evaluation results of SSIM.

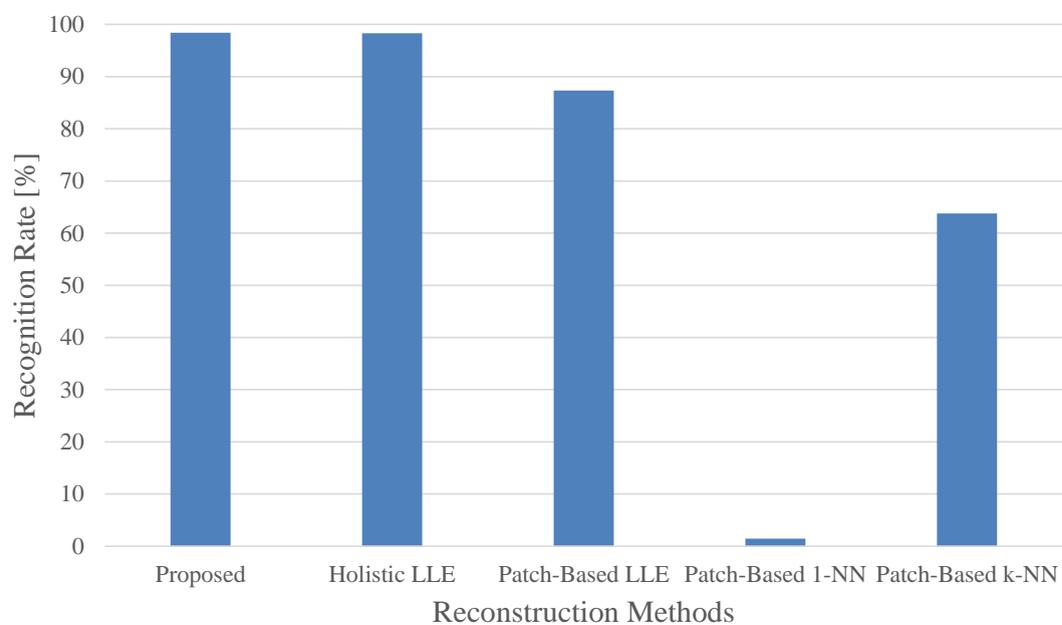


FIGURE 4.8: Evaluation results of Face Recognition.

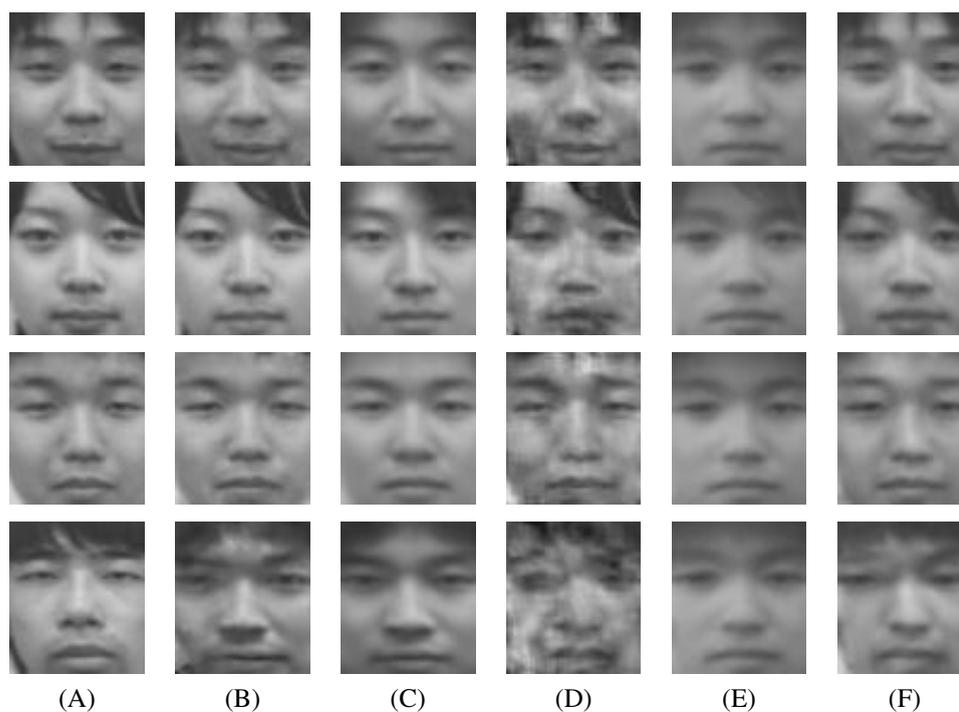


FIGURE 4.9: Examples of reconstructed images by various methods. Each row shows an individual person and the columns contain: (A) Ground-truth images; (B) Reconstructed images of the proposed method; (C) Reconstructed images of Holistic LLE (Locally Linear Embedding); (D) Reconstructed images of Patch-Based LLE; (E) Reconstructed images of Patch-Based 1-NN (Nearest Neighbor); (F) Reconstructed images of Patch-Based  $k$ -NN.

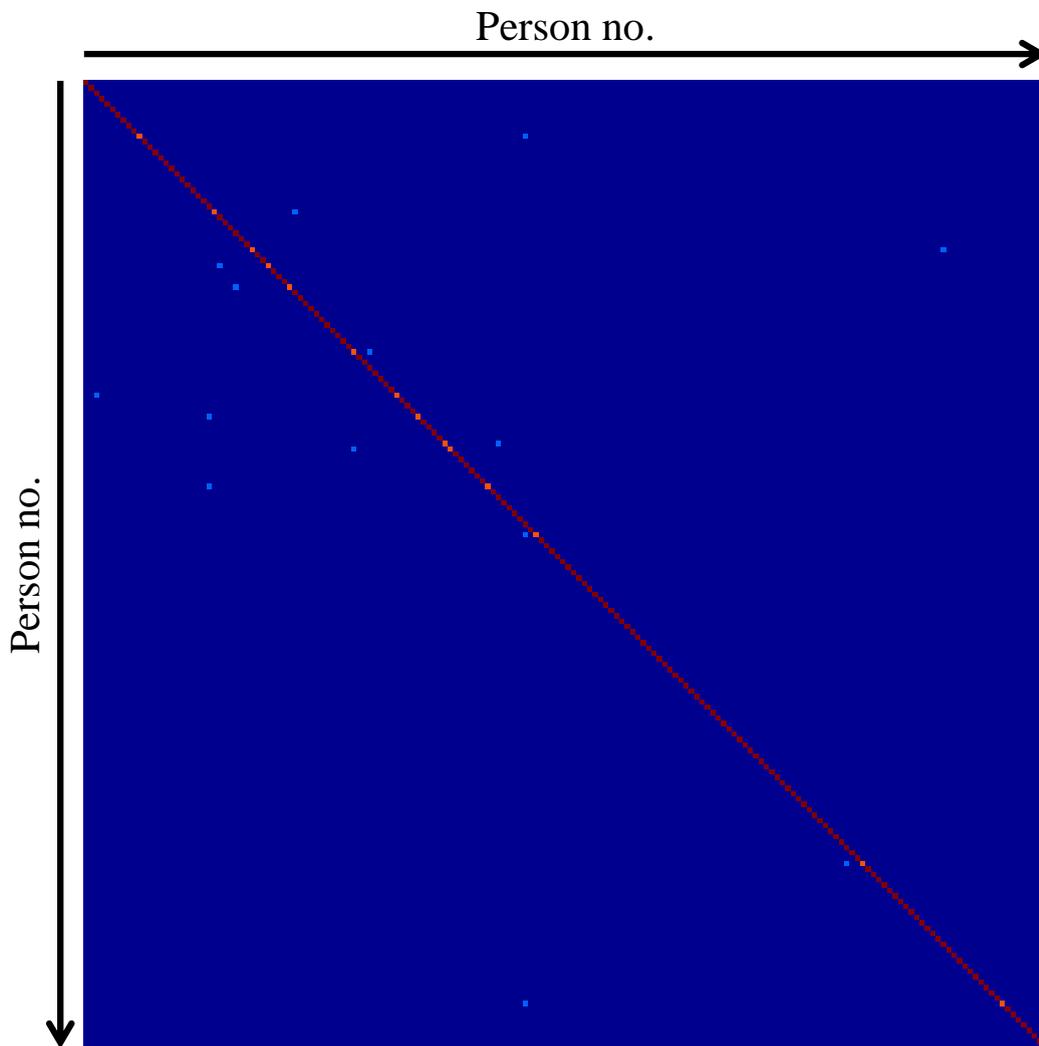


FIGURE 4.10: Heat map representation of the confusion matrix of the recognition evaluation. It goes from dark blue to dark red, where the representations of higher values are close to dark red.

We can see from the results that the proposed method outperformed other comparative methods in both quality evaluation measures. Since the proposed method employs both full image reconstruction to create a basic face image in the Global Reconstruction step and refine it through the reconstruction of patches in the Local Refinement step, it can be concluded that this combination plays an important role in this quality evaluation. To reinforce this notion, we can also observe the quality from the actual reconstruction results shown in Figure 4.9. The proposed method produced results that resembled closely to its ground-truth, unlike other comparative methods which faced problems in doing so. The Holistic LLE produced face images

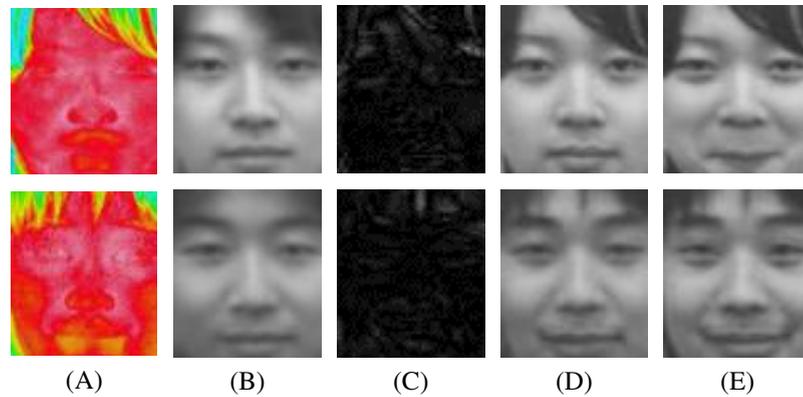


FIGURE 4.11: Visual examples in various steps of the proposed method. Each row shows an individual person and the columns contain: (A) Thermal infrared input images; (B) Globally reconstructed images; (C) Residual images (Normalized to 0–1); (D) Fully reconstructed images; (E) Ground-truth images.

that lacked personality and details. Because of this, the reconstructed face can easily be mistaken with someone else. The Patch-Based LLE produced face images relatively better than those produced by the Holistic LLE; They look more similar to the ground-truth. However, there are more artifacts in them and the reconstructed faces are quite unnatural. The Patch-Based 1-NN method produced face images that lack details the most out of all the methods. It seems that the usage of only one visible patch for the reconstruction significantly affected the results. The Patch-Based  $k$ -NN method produced relatively detailed face images even though they are not as good as those produced by the proposed method.

In the face recognition evaluation, the proposed method also outperformed all other comparative methods. Although the proposed method showed the best performance, the recognition rate of the Holistic LLE method was also good. This is because EigenFace was used as the recognition method, where only the prominent features of the face actually affected the recognition results. As a result, the details of the reconstructed face images did not influence the recognition rate as much, rendering the details reintroduced by the Local Refinement step less advantageous. As for the patch-based methods, they were all outperformed by the proposed method and the Holistic LLE method. Among all methods, the Patch-Based 1-NN performed the worst.

As we can see in the last row of Figure 4.9, some mis-reconstruction occurred. As humans, we can see that the reconstructed face and the ground-truth are not the same person. As a matter of fact, all methods struggled to correctly reconstruct this person. The most likely reason for this would be that there were faces whose thermal patterns were similar to this person in the dataset, causing LLE to fail in finding the correct neighbors. Another possible reason is correlated to the geometric relations between the nearest neighbors of the thermal and its visible counterpart. Alternatively, the existence of hair on the upper part of the image, might have also hindered the reconstruction process.

It can be concluded that the proposed method takes the best of both holistic only and patch-based only approaches with satisfying results across two evaluation criteria; quality and recognition.

## 4.4 Discussion

In this section, additional experimental results are provided for a more detailed discussion. This includes the additional experiments performed to obtain the optimal set of parameters for the proposed method and another set of experiments to evaluate the proposed method in a more difficult situation.

### 4.4.1 Finding the Optimal Set of Parameters

In this experiment, two adjustable parameters are set; patch size for the Local Refinement step and the number of nearest neighbors. The options for patch size were  $5 \times 5$ ,  $7 \times 7$ , and  $9 \times 9$  pixels. The number of neighbors used by the LLE method was selected from either 5, 15, or 30. Table 4.3 shows the results with the combinations of the parameters. Bar graph representations are also shown in Figures 4.12, 4.13 and 4.14 for the PSNR, SSIM, and recognition rate, respectively. Based on these results, as mentioned previously, the selected set of parameters was five nearest neighbors and

TABLE 4.3: Evaluation of various patch sizes with various numbers of neighbors. SE stands for the standard error, and is represented in the bracket.

# of Neighbors	Patch Size [Pixels]	PSNR (SE) [dB]	SSIM (SE)	Recog. Rate [%]
5	5 × 5	31.40 (3.20)	0.93 (0.05)	98.33
	7 × 7	32.57 (3.57)	0.94 (0.05)	98.11
	9 × 9	<b>33.11</b> (3.69)	<b>0.95</b> (0.05)	<b>98.44</b>
15	5 × 5	29.88 (2.71)	0.90 (0.06)	97.78
	7 × 7	31.66 (3.28)	0.93 (0.05)	97.67
	9 × 9	32.41 (3.50)	0.94 (0.05)	97.78
30	5 × 5	29.03 (2.48)	0.89 (0.05)	97.56
	7 × 7	30.31 (2.88)	0.92 (0.05)	97.56
	9 × 9	31.31 (3.20)	0.93 (0.05)	97.67

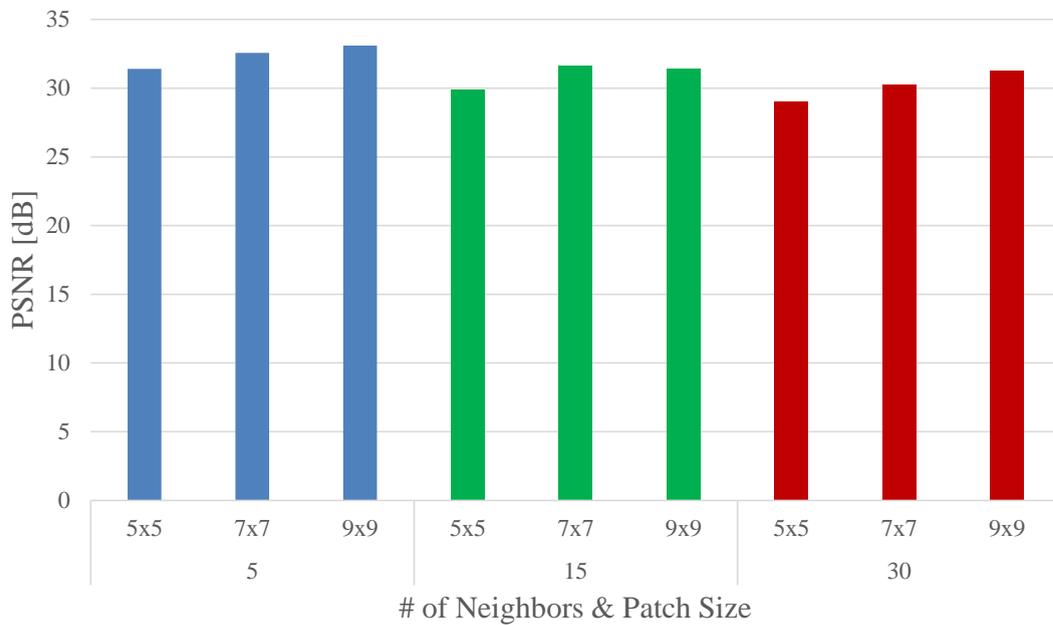


FIGURE 4.12: Evaluation results of PSNR.

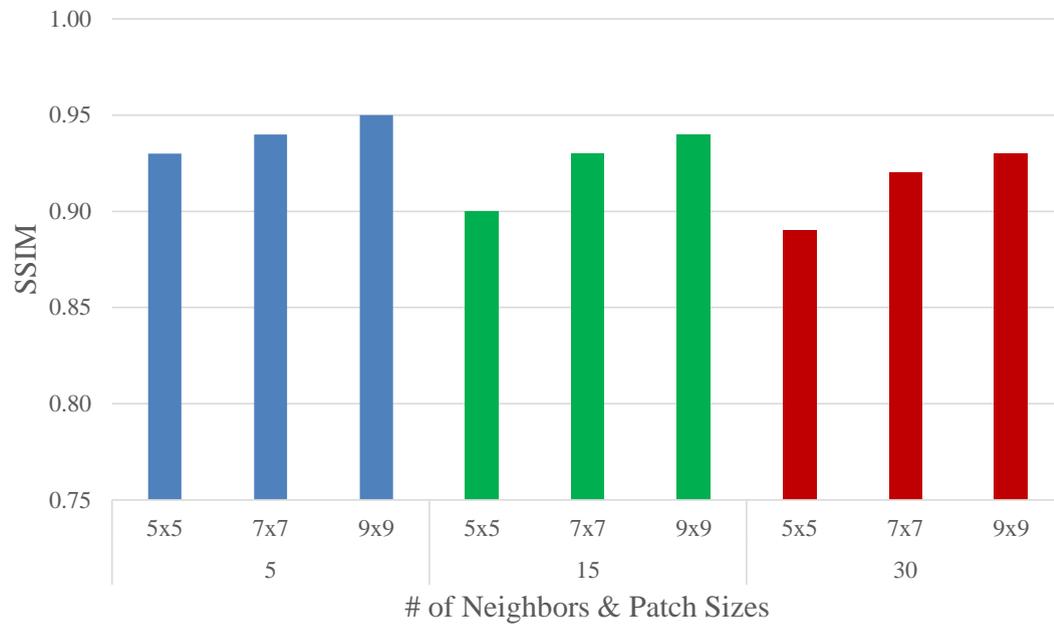


FIGURE 4.13: Evaluation results of SSIM.

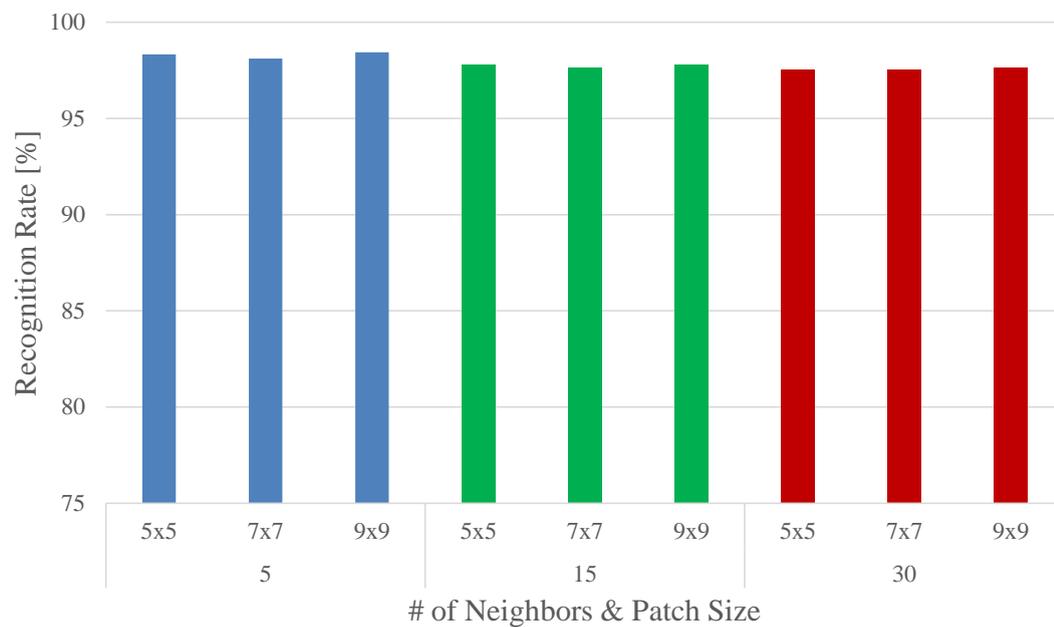


FIGURE 4.14: Evaluation results of Face Recognition.

9 × 9 pixels patch size. In all evaluation methods, this set of parameters achieved the highest results, outperforming other sets of parameters.

#### 4.4.2 Face Image Reconstruction of Unknown Persons

In Section 4.3, the experiment was conducted with the assumption that variations of the face to be reconstructed is included in the training data. This means that the proposed method is relatively prepared for the reconstruction. For further discussion, experiments were performed where this condition is not met, and tried to reconstruct the face image of an unknown person. This case is more difficult and challenging, because the proposed method does not learn the relationship between the thermal patterns and the visible information of the person beforehand.

Two experiments were performed in this case. The first experiment was performed to see the performance of the reconstruction. The second experiment was performed to see the effect of increasing the size of training dataset.

##### 4.4.2.1 Performance Evaluation

This experiment was conducted with the motivation to investigate the feasibility of reconstructing a face image which was not included in the training phase. Since the proposed method does not learn the relationship between the thermal patterns and the visible information of the face to be reconstructed, it needs to utilize the relationship learned from other subjects in the training data.

As this case is fundamentally different with the previous experiment, the division of the dataset also differs. First, twenty subjects were taken from the available 180 to be used later in the recognition evaluation. The remaining 160 subjects were then divided into sixteen groups with ten subjects each. Cross-validation was performed on these sixteen groups, training fifteen groups and testing the other. After the reconstruction of a test group was done, the ground-truths of the said group and the

TABLE 4.4: Comparison of the proposed method with the holistic method and various patch-based methods (Number of neighbors: 5, Patch size:  $9 \times 9$  pixels). SE stands for the standard error, and is represented in the bracket.

Method	PSNR (SE) [dB]	SSIM (SE)	Recog. Rate [%]
Proposed	19.36 (3.11)	<b>0.70</b> (0.13)	12.25
Holistic LLE	19.39 (3.11)	<b>0.70</b> (0.13)	11.38
Patch-Based LLE	<b>19.46</b> (2.40)	0.69 (0.09)	<b>23.13</b>
Patch-Based 1-NN	18.05 (1.64)	0.65 (0.07)	3.13
Patch-Based $k$ -NN	19.26 (2.41)	0.69 (0.09)	19.88

excluded twenty subjects were grouped together for the recognition evaluation. This scheme keeps the difficulty of the recognition process high, and avoids trained data points to be used in the recognition evaluation because that would raise the possibility of misclassification.

This first experiment is similar to the experiment in Section 4.3, with the same comparative methods and evaluation criteria. The results of this experiment are shown in Table 4.4. Bar graph representations are also shown in Figures 4.15, 4.16, and 4.17 for the PSNR, SSIM, and recognition rate, respectively. The results show that the reconstruction of unknown persons is currently not feasible. Improvement of the proposed method is necessary to increase its performance. For example, instead of learning the relationship between the visible and the thermal infrared spectra by assuming it is linear, a non-linear relationship can be studied by using a kernel. Kernel Canonical Correlation Analysis [117–121] is a good option for immediate future work. There is also a deep learning approach to CCA, appropriately called as Deep Canonical Correlation Analysis (DCCA) [122].

Another way to possibly improve the performance is by increasing the size of the training dataset, which effects the variety of faces available for the reconstruction. The next experiment is related to this point, and will be discussed in depth in Section 4.4.2.2.

The difference of quality evaluation between all of the methods were relatively negligible, except for the Patch-Based 1-NN with an abysmal score. This is the exact

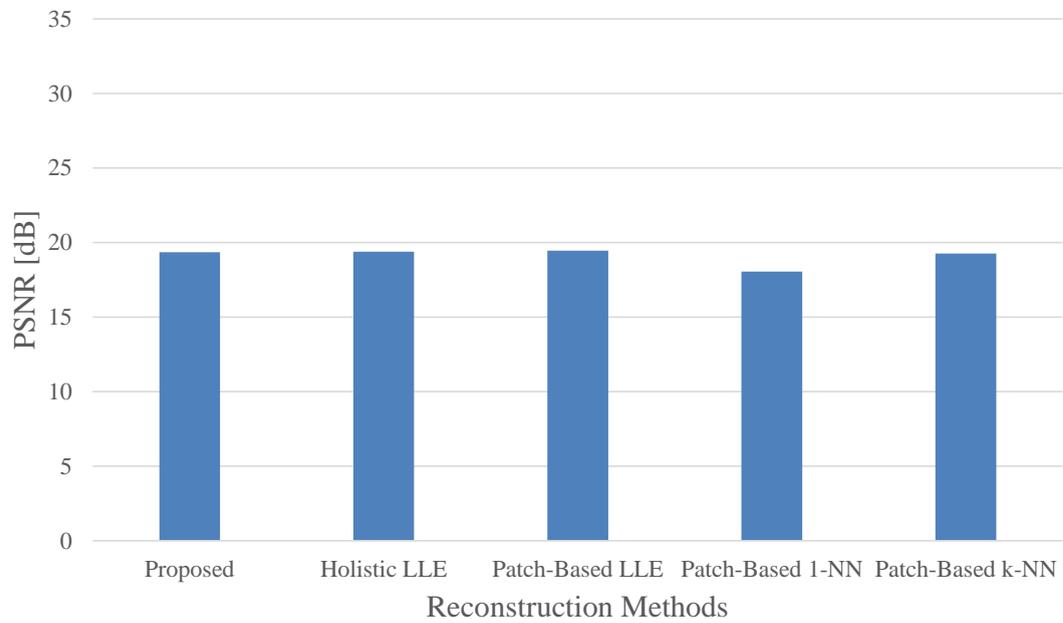


FIGURE 4.15: Evaluation results of PSNR.

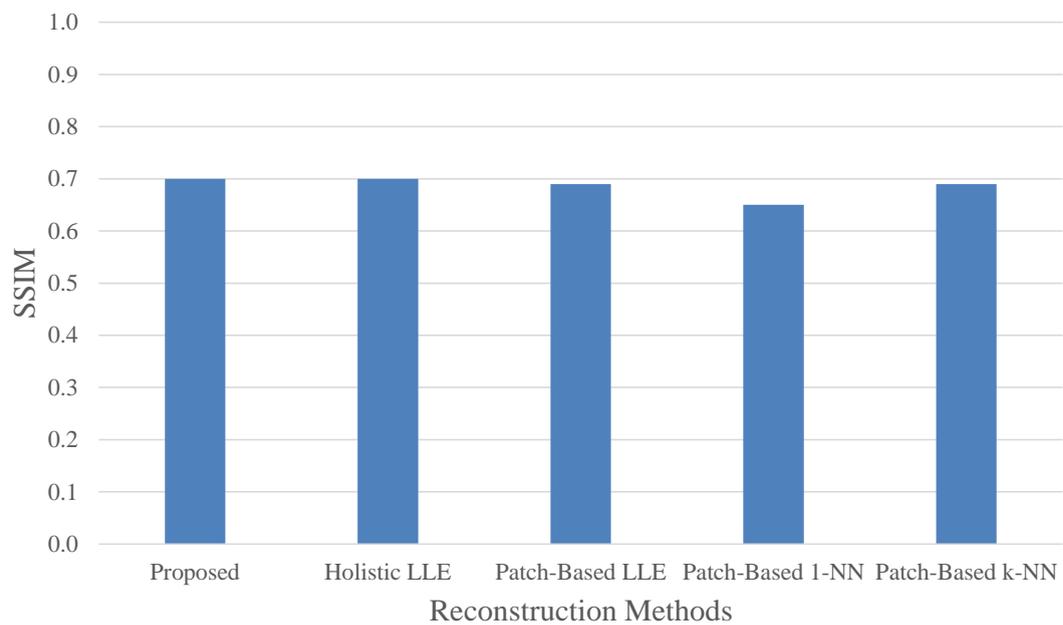


FIGURE 4.16: Evaluation results of SSIM.

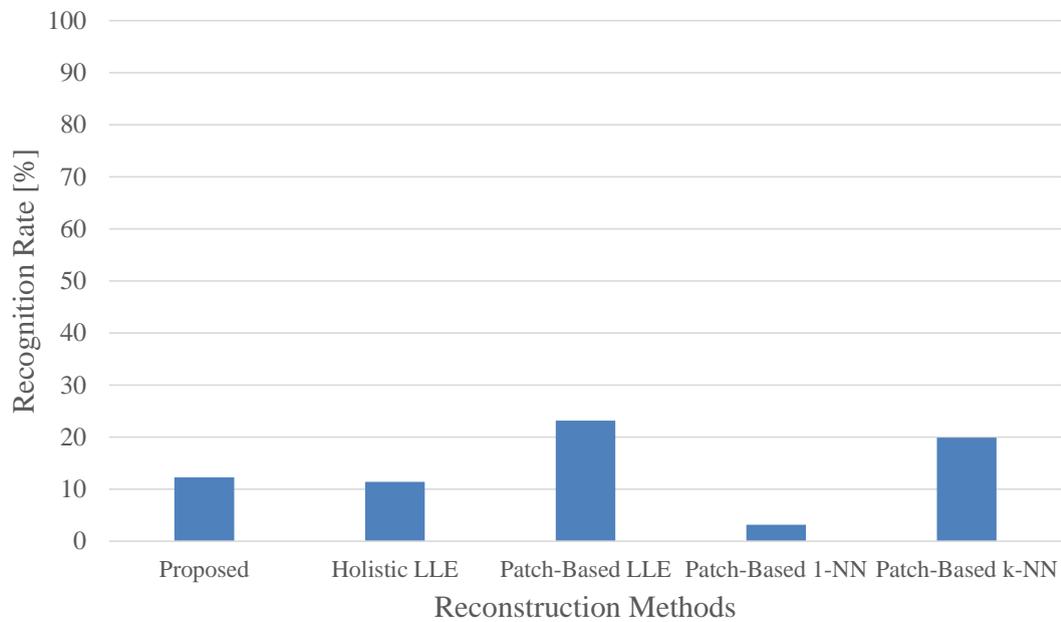


FIGURE 4.17: Evaluation results of Face Recognition.

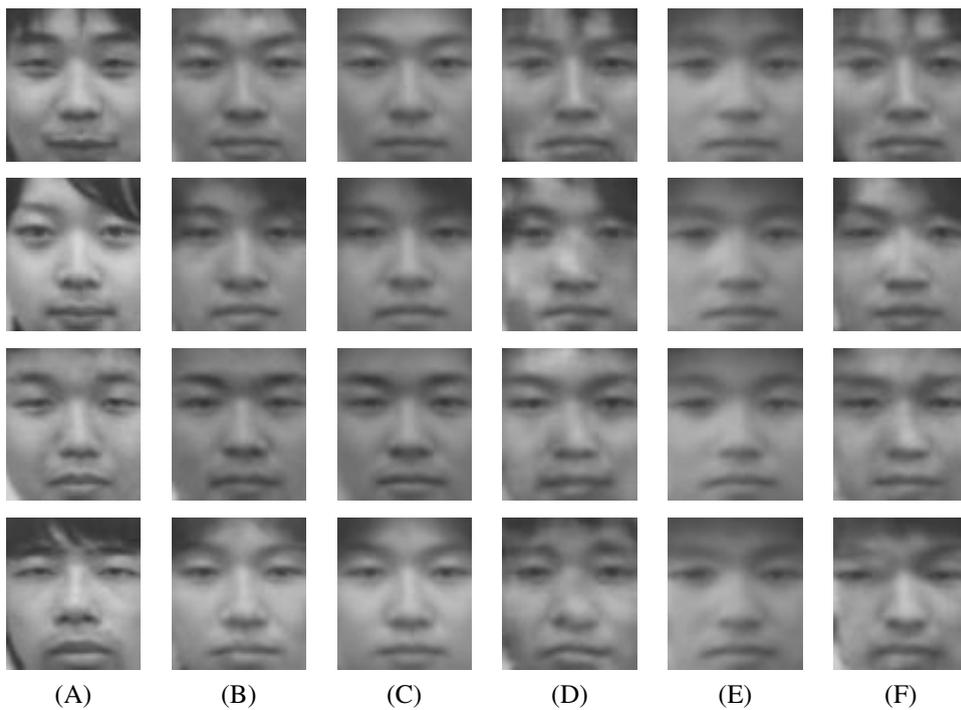


FIGURE 4.18: Examples of reconstructed images by various methods of unknown persons. Each row shows an individual person and the columns contain: (A) Ground-truth images; (B) Reconstructed images of the proposed method; (C) Reconstructed images of Holistic LLE (Locally Linear Embedding); (D) Reconstructed images of Patch-Based LLE; (E) Reconstructed images of Patch-Based 1-NN (Nearest Neighbor); (F) Reconstructed images of Patch-Based  $k$ -NN.

opposite to the trend observed in the experiment conducted in Section 4.4.2.2, where the proposed method outperformed all of the comparative methods. For the face recognition evaluation, the Patch-Based LLE performed the best. In this case, the most likely reason is the reconstruction of the whole face image encounters problems. When the base face image from the results of the Global Reconstruction step is not good enough or far from the ground-truth, reintroducing details with the Local Refinement step cannot improve the results much. On the other hand, the Patch-Based LLE does not make use of the base image and reconstructs the face directly in small patches and combines them. This method reduces the possibility of misreconstruction on the whole face image which is more detrimental to the recognition performance.

Figure 4.18 shows the actual reconstruction results. We can see that all of the methods struggled to reconstruct the faces satisfactorily. The reconstruction results of the proposed method can be seen as an enhanced version of the Holistic LLE method. All variations of the patch-based methods produced results that looked unnatural in multiple areas of the face. Conclusion can be made from this experiment that the face image reconstruction of an unknown person is a difficult task.

#### 4.4.2.2 Various Sizes of Training Dataset

The second experiment was conducted to investigate the effect of the size of training dataset to the reconstruction. This is related to the fact that the proposed method does not know how an unknown thermal face would look like in the visible spectrum. By increasing the size of the training dataset, the available face variations for the reconstruction grows. This can increase the possibility of the proposed method in finding a closer match in the LLE method.

To conduct the experiment, the dataset was divided as follows: first, twenty subjects were separated to be used later in the recognition evaluation, just like in the experiment in Section 4.3. Next, additional ten subjects were chosen to be used solely for the testing. Five subsets were then created from the rest of the data with different

TABLE 4.5: Evaluation of different numbers of training data (Number of neighbors: 5, Patch size:  $9 \times 9$  pixels). SE stands for the standard error, and is represented in the bracket.

# of Subjects	PSNR (SE) [dB]	SSIM (SE)	Recog. Rate [%]
30	18.74 (2.82)	0.72 (0.09)	4.00
60	19.63 (2.68)	0.74 (0.07)	6.00
90	19.64 (2.89)	0.72 (0.10)	10.00
120	19.80 (2.84)	0.73 (0.10)	6.00
150	20.26 (2.52)	0.73 (0.09)	14.00

numbers of subjects in it. The reconstructed test data were combined with the twenty separated subjects for the recognition evaluation. It is important to note that this experimental setup always used the same ten subjects for the testing, but the numbers of subjects in the training data differed.

The results of this experiment are shown in Table 4.5. Bar graph representations are also shown in Figures 4.19, 4.20, and 4.21 for the PSNR, SSIM, and recognition rate, respectively. While the score is low overall, the linear extrapolations from the results show an increasing tendency. It can be concluded that the existence of more variations in the training data helped the reconstruction process and a conjecture can be made that a very large amount of training data may produce a satisfactory result in reconstructing the face image of an unknown person.

The actual reconstructed face images by the proposed method are shown in Figure 4.22. Although the results do not resemble the ground-truths, changes of the reconstructed face as the number of the training increases can be seen. The most significant changes can be observed when the training data increased from 30 subjects to 60 subjects. Only small changes could be seen beyond that.

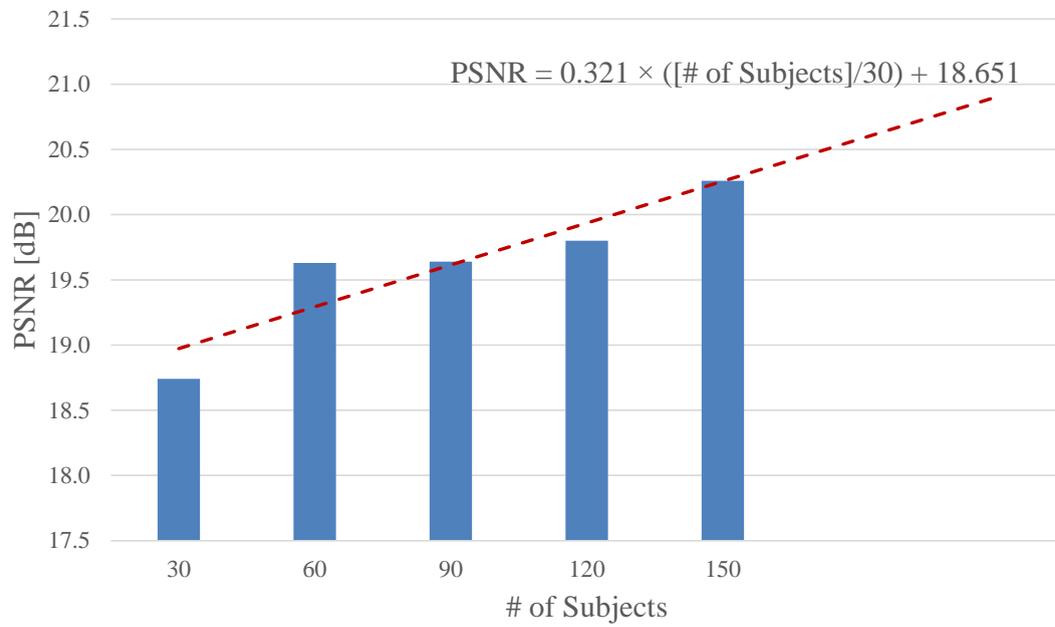


FIGURE 4.19: Evaluation results of PSNR. The PSNR was linearly extrapolated (red dashed line).

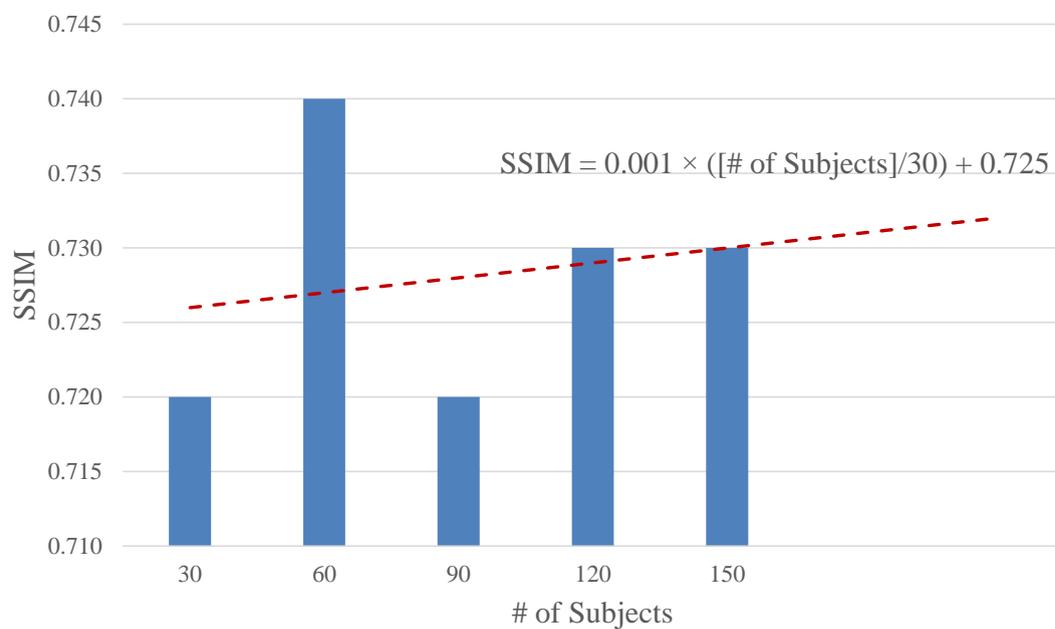


FIGURE 4.20: Evaluation results of SSIM. The SSIM was linearly extrapolated (red dashed line).

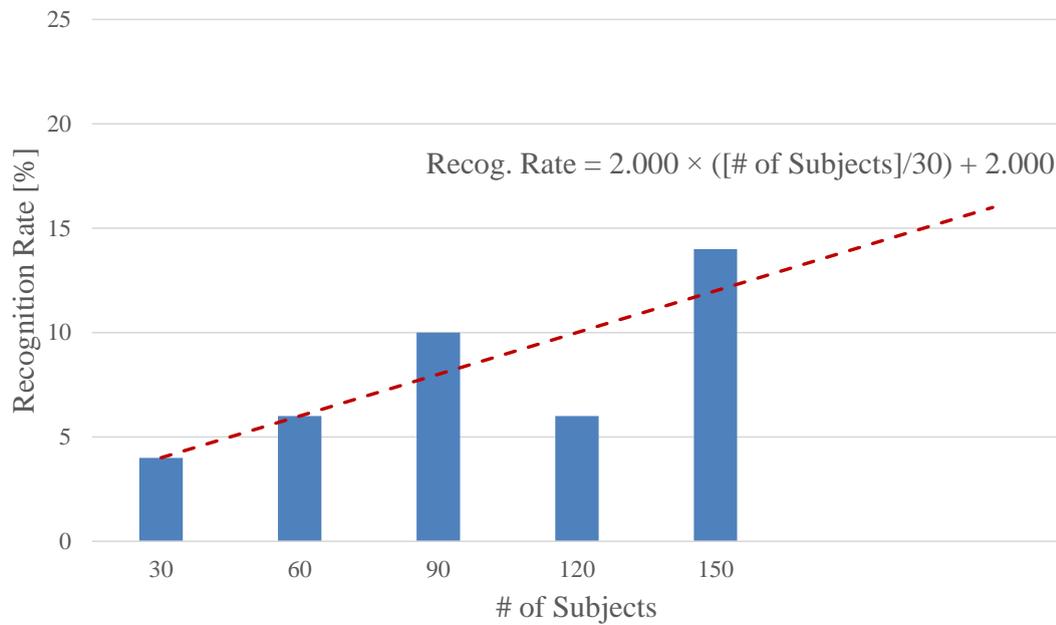


FIGURE 4.21: Evaluation results of Face Recognition. The recognition rate was linearly extrapolated (red dashed line).

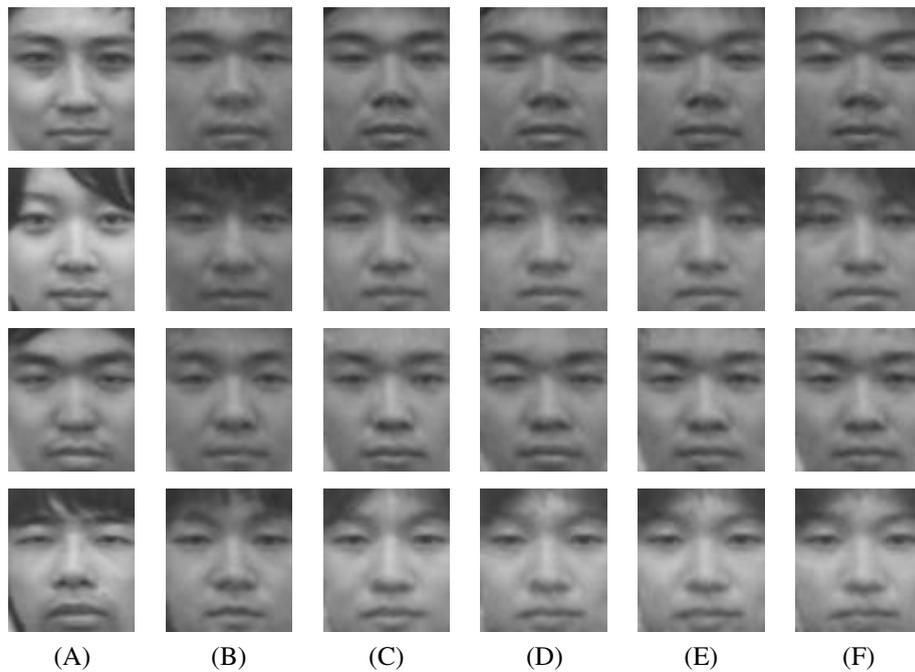


FIGURE 4.22: Examples of reconstructed images with various numbers of training data. Each row shows an individual person and the columns contain: (A) Ground-truth images; (B) Reconstructed images from 30 subjects' training data; (C) Reconstructed images from 60 subjects' training data; (D) Reconstructed images from 90 subjects' training data; (E) Reconstructed images from 120 subjects' training data; (F) Reconstructed images from 150 subjects' training data.

## 4.5 Summary

This chapter addressed the answer to the first research question in this thesis, which was “What does the person look like in the visible spectrum?” This chapter tackled a face image reconstruction task from the thermal infrared spectrum to the visible spectrum to answer this question. In order to achieve this, a two-step reconstruction method was proposed. The first step was referred to as the Global Reconstruction, where the reconstruction was performed on the whole image. The second one was referred to as the Local Refinement, where patches of the image were reconstructed to refine the results of the previous step. The proposed method utilized CCA in the training phase in order to visually understand the relationship between the thermal and the visible images. For the reconstruction of the face image, the proposed method exploited the aforementioned relationship with LLE to reconstruct the face image.

Experiments were conducted to evaluate the reconstruction capability of the proposed method. The proposed method showed high performance in all evaluations and outperformed other comparative methods. A discussion was also provided on the reconstruction of an unknown person’s thermal face image. The proposed method together with comparative methods struggled to perform this kind of reconstruction because there was no information of the face in the visible spectrum. This proves the difficulty of the task and warrants further research in the field.



## Chapter 5

# Semantic Understanding from Thermal Infrared Image

There are two research topics in this thesis, each of which is related to the understanding of human images in the thermal infrared spectrum. This chapter presents the second topic; Semantic understanding from thermal infrared image. This chapter is structured as follows. Section 5.1 describes the general background of the wearable attribute recognition research and its relation to the theme of the thesis; Thermal infrared image understanding. Section 5.2 explains Robust Principal Component Analysis (RPCA), the core method used in this research, and the modification made to it. Section 5.3 elaborates two versions of the framework proposed to gain semantic understanding of the human image in the thermal infrared spectrum. The approach to the problem is by decomposing the human image into multiple components using RPCA. After the decomposition is performed, it is possible to recognize which wearable attributes are present in the human image. Next, the experiments along with its results and analysis are described in Section 5.4. The chapter is closed with the summary in Section 5.5.

## 5.1 Background

In Chapter 1, the second research question in this thesis was discussed. The question was “What is the person wearing?” The research topic corresponding to it proposes the recognition of wearable attributes to answer this question. In places such as bank, ATM areas, military outposts, the presence of a person with wearable attributes over the face itself is a security risk, because the face is occluded by the attributes. Therefore, the information of what wearable attributes a person is wearing is equally important to avoid a compromise in security. The recognition task can then be expanded for searching people based on attributes they are wearing or to warn security personnels about the usage of a certain attribute.

In order to perform the recognition, the system is required to have a certain understanding regarding human images in the thermal infrared spectrum. In this thesis, the understanding is referred to as semantic understanding. The research in this chapter gains this understanding by decomposing the human image using RPCA which is described in Section 5.2. The details of the two proposed frameworks in this research are provided in Section 5.3.

## 5.2 Robust Principal Component Analysis (RPCA)

This section starts by introducing the basis of the proposed method, Robust Principal Component Analysis (RPCA) and the innovative use of RPCA to recognize wearable attributes. The incorporation of the Probability Map (PM) as a prior knowledge to achieve a more focused decomposition is described next. This is a novel addition to the RPCA method.

### 5.2.1 RPCA via Principal Component Pursuit (PCP)

Principal Component Analysis (PCA) is potentially one of the most commonly used statistical tools for dimensional reduction and data analysis. However, PCA encounters problems when dealing with grossly corrupted data. This is because when there are many outliers and noisy observations in the data, the estimated principal components might not be correct. In these cases, PCA cannot transform the data into a new coordinate system where the coordinates correspond to the principal components.

Robust Principal Component Analysis (RPCA or Robust PCA) is a modification on PCA made to be robust towards corrupted data. Introduced by Candes et al. [18], an idealized version of RPCA aims to recover a low-rank matrix from the corrupted data regardless of the magnitude of the noise. In simpler terms, Robust PCA handles the corrupted/noisy observations by separating them from the data. The idealized version of RPCA decomposes a collection of observations  $\mathbf{M}$  to a low-rank matrix  $\mathbf{L}$  and a sparse matrix  $\mathbf{S}$ , as follows:

$$\mathbf{M} = \mathbf{L} + \mathbf{S}, \quad (5.1)$$

where sparse matrix  $\mathbf{S}$  contains the noisy/corrupted part of the observation in the data. The magnitude of elements of  $\mathbf{S}$  can be arbitrarily large. For the purpose of the decomposition, the RPCA is utilized by the proposed method with the majority-minority idea mentioned in Section 1.3.2. Therefore, the desired output is the sparse representation of the input contained in sparse matrix  $\mathbf{S}$ .

There are various techniques to solve the RPCA problem. Principal Component Pursuit (PCP) [18] is still one of the best techniques among its peers [123, 124] even though some adaptations have been proposed for it such as Stable PCP [125], Local PCP [126], and others [127–133].

Let us assume a data matrix  $\mathbf{M}$  where the observations are represented in the form of a column vector. PCP solves the following optimization problem:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{s. t.} \quad \mathbf{L} + \mathbf{S} = \mathbf{M}, \quad (5.2)$$

where  $\|\cdot\|_*$  denotes the nuclear norm, which is the sum of the singular values given a matrix.  $\|\cdot\|_1$  is the  $l_1$ -norm with the condition that the matrix is treated as a vector, and  $\lambda$  is a balance parameter. The value of  $\lambda$  obeys the rule of thumb as follows:

$$\lambda = \frac{1}{\sqrt{\max(m, n)}}, \quad (5.3)$$

where  $m$  and  $n$  represent the number of rows (dimensions) and columns (observations) of matrix  $\mathbf{M}$ , respectively. In most cases, the value of  $\lambda$  does not need to be adjusted except when prior knowledge is available. PCP is able to obtain the low-rank and the sparse matrices of a data matrix under minimal assumptions. These assumptions are that the low-rank matrix  $\mathbf{L}$  is not sparse, the sparse matrix  $\mathbf{S}$  is not low-rank, and the value of  $\lambda$  follows the rule of thumb in Equation (5.3).

### 5.2.2 RPCA via Principal Component Pursuit (PCP) with Probability Map

The main idea behind the extension of PCP is to extract attribute information on a smaller region. This can be achieved by the incorporation of prior knowledge, which in this case, is a region where an attribute is likely to be present. As an example, to recognize the “glasses” attribute, the only region that needs to be checked is the region around the eyes. Even if there are “glasses” in another area of the body, for example, in a breast pocket, it is not necessary to recognize them.

Based on this idea, this research proposes RPCA via Probability Map - Principal Component Pursuit (PM-PCP). The Probability Map is a representation of how likely elements of sparse component will be present in a given matrix. Using the original RPCA problem in Equation (5.1), the decomposition problem with Probability Map can be described as follows:

$$\mathbf{M} = \mathbf{L} + \mathbf{S} = \mathbf{L}' + \mathbf{P} \circ \mathbf{S}, \quad (5.4)$$

where  $\mathbf{P}$  is the Probability Map, and  $\mathbf{P} \circ \mathbf{S}$  denotes an element-wise multiplication between matrices  $\mathbf{P}$  and  $\mathbf{S}$ .

In this extension of PCP, it is important to highlight the Probability Map. The Probability Map  $\mathbf{P}$  is used as a weight of the sparse matrix in the decomposition, represented by  $\mathbf{P} \circ \mathbf{S}$ . Since it extracts sparse elements only on the region where elements of  $\mathbf{P}$  are greater than 0, other sparse elements are not extracted and left in  $\mathbf{L}$ . Therefore, the RPCA problem can now be rewritten as follows:

$$\mathbf{M} = \mathbf{L} + (\mathbf{1} - \mathbf{P}) \circ \mathbf{S} + \mathbf{P} \circ \mathbf{S}, \quad (5.5)$$

where there are two occurrences of the sparse matrix  $\mathbf{S}$ . The first one  $(\mathbf{1} - \mathbf{P}) \circ \mathbf{S}$  represents sparse entries excluding sparse elements not extracted from  $\mathbf{M}$  while the second one  $\mathbf{P} \circ \mathbf{S}$  contains the said sparse elements. For the purpose of decomposition, only the second sparse matrix term is needed. Therefore, the problem can then be changed to the following:

$$\mathbf{M} = \mathbf{L}' + \mathbf{P} \circ \mathbf{S}, \quad (5.6)$$

where

$$\mathbf{L}' = \mathbf{L} + (\mathbf{1} - \mathbf{P}) \circ \mathbf{S}. \quad (5.7)$$

In this case,  $\mathbf{L}'$  contains both the low-rank representation and noise from the sparse matrix, which can be discarded because it is not necessary in order to achieve the end goal. On the other hand,  $\mathbf{P} \circ \mathbf{S}$  contains the sparse elements from a specific region based on the Probability Map  $\mathbf{P}$ , the output desired from the decomposition.

As we can see, the assumption made earlier in Equation (5.4) holds true. The original minimization problem by PCP (Equation (5.2)) is then modified to incorporate the weighting by the Probability Map as follows:

$$\min_{\mathbf{L}', \mathbf{P} \circ \mathbf{S}} \|\mathbf{L}'\|_* + \lambda \|\mathbf{P} \circ \mathbf{S}\|_1 \quad \text{s. t.} \quad \mathbf{L}' + \mathbf{P} \circ \mathbf{S} = \mathbf{M}. \quad (5.8)$$

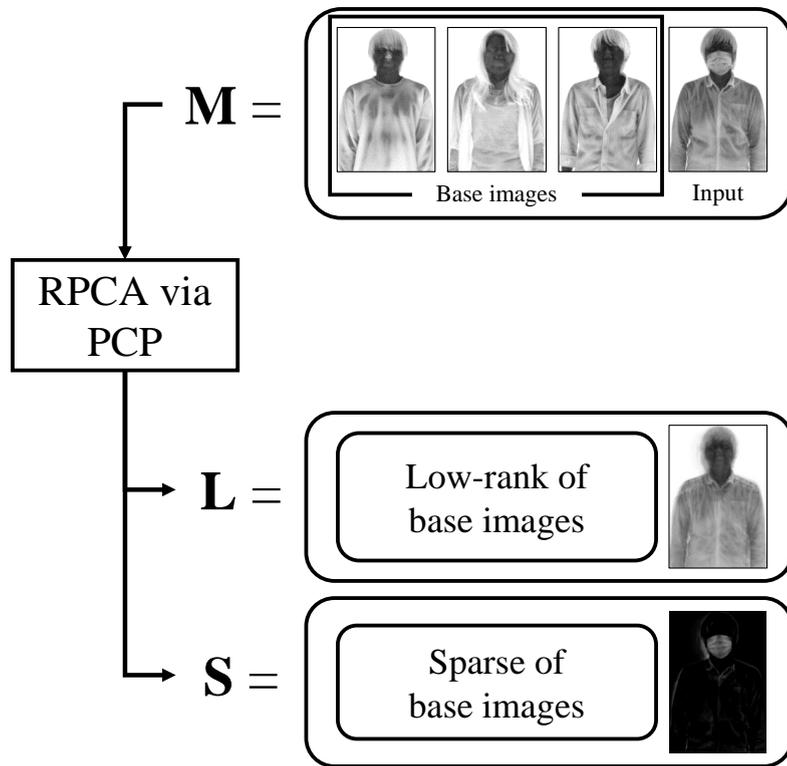
It is important to note that although the minimization problem is modified due to the incorporation of Probability Map, the original algorithm used for solving the PCP can still be used with only a slight modification. This is because the PM is constant, hence it does not affect the convergence. The only difference is that there is an element-wise multiplication on the sparse matrix by the Probability Map.

### 5.3 Recognition Framework

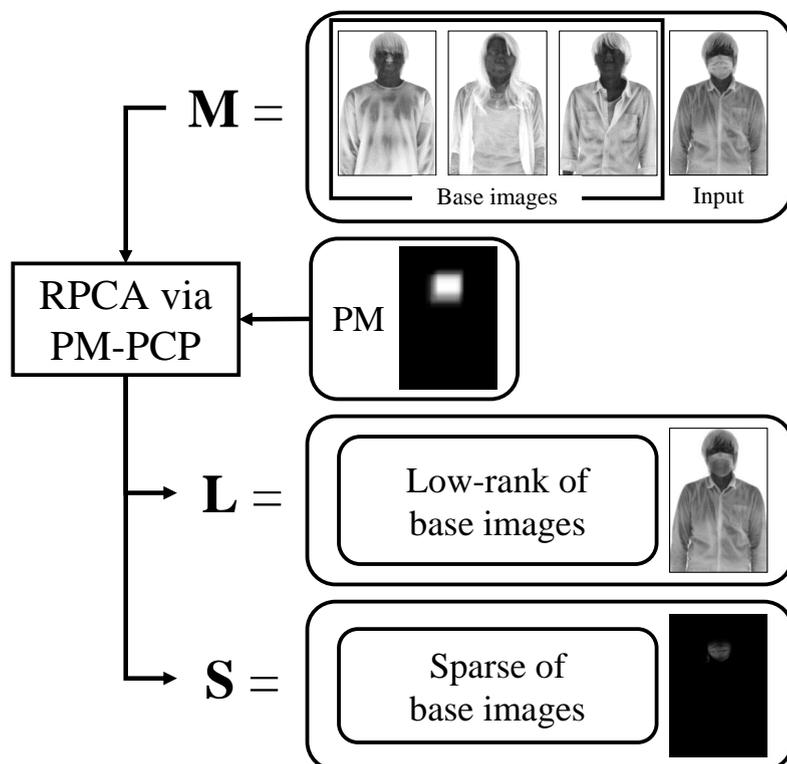
This section describes the recognition framework in a more detailed fashion. There are two versions of the framework explained in this section; the first one makes use of RPCA via PCP and the second one RPCA via PM-PCP.

As mentioned previously, the RPCA is utilized by the proposed method with the majority-minority idea. The proposed method achieves this by controlling the data matrix  $\mathbf{M}$ , effectively exploiting RPCA's capability of decomposition. The proposed method only uses the sparse matrix output of RPCA, since it contains the attributes and represents the minority. Figure 5.1 shows how the RPCA via PCP and PM-PCP are used to perform decomposition in the framework.

It is important to note that the thermal infrared images utilized in the process uses the “hotblack” color scheme. The “hotblack” scheme is a monochrome color scheme that considers the temperature captured in the pixel, where the hotter the temperature, the closer the pixel value is to zero.



(A) RPCA via PCP



(B) RPCA via PM-PCP

FIGURE 5.1: Usage of RPCA for decomposition.

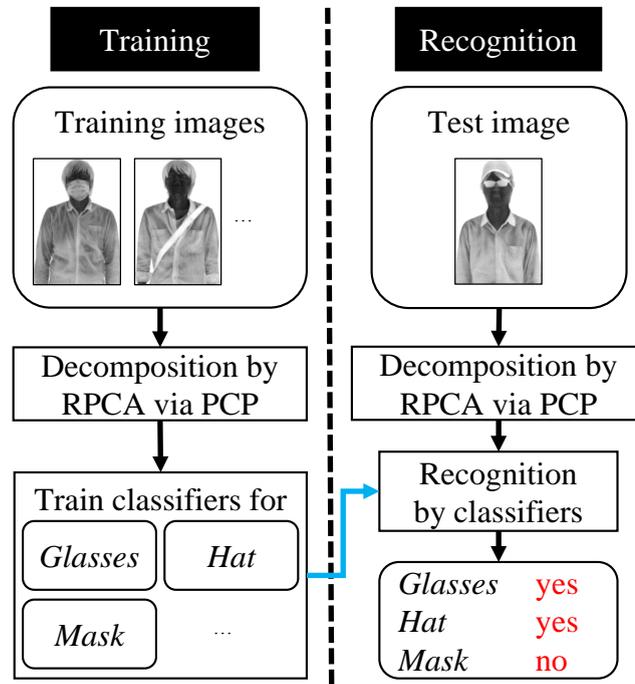


FIGURE 5.2: Process flow of the recognition framework by RPCA via PCP.

### 5.3.1 Recognition Framework with RPCA via PCP

The overall process flow of the framework can be seen in Figure 5.2, which is divided into training and recognition phases. The decomposition process is performed on each phase, followed by the training/recognition of the classifier.

First, from the available dataset, images of people with no wearable attributes are chosen as base data  $\mathbf{B} = [\mathbf{b}_1 \mathbf{b}_2 \cdots \mathbf{b}_H]$  with  $H$  observations. This data is used in both training and recognition phases, to serve as the majority in the majority-minority idea of decomposition.

In the training phase, the training data is represented as  $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \cdots \mathbf{w}_J]$  where  $J$  is the number of training data. It is important to note that base data and training data do not intersect. For the decomposition idea of majority-minority, one observation from the training data  $\mathbf{W}$  is chosen to be the minority. Then, PCP is used to solve the RPCA problem upon these datasets as follows:

$$\begin{bmatrix} \mathbf{B} & \mathbf{w}_j \end{bmatrix} = \mathbf{M}_j = \mathbf{L}_j + \mathbf{S}_j, \quad (5.9)$$

where

$$\mathbf{S}_j = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_H & \mathbf{x}_j \end{bmatrix}. \quad (5.10)$$

In Equation (5.10),  $\mathbf{x}_j$  is an entry of sparse matrix  $\mathbf{S}_j$  that corresponds to  $\mathbf{w}_j$ , where  $j = 1, 2, \dots, J$ . After PCP is performed  $Q$  times, the training data is arranged as  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_j]$  and further used for training by a classifier. Any classifier can be used for this framework, since the proposed method has both positive and negative data. The training phase is concluded when the classifier is trained for the attribute.

In the recognition phase, the assumption is that wearable attributes were to be recognized from a new thermal infrared data  $\mathbf{t}$ . Using the base data  $\mathbf{B}$  as the majority and the thermal infrared data  $\mathbf{t}$  as the minority, the decomposition can then be performed according to plan. The optimization problem that is needed to be solved is the following:

$$\begin{bmatrix} \mathbf{B} & \mathbf{t} \end{bmatrix} = \mathbf{M}_t = \mathbf{L}_t + \mathbf{S}_t, \quad (5.11)$$

where

$$\mathbf{S}_t = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_H & \mathbf{x}_t \end{bmatrix}. \quad (5.12)$$

Similar to the training phase, in Equation (5.12),  $\mathbf{x}_t$  is an entry of sparse matrix  $\mathbf{S}_t$  that corresponds with  $\mathbf{t}$ . The component  $\mathbf{x}_t$  is then used for recognition by the classifier in the next step. In the recognition phase, test data  $\mathbf{x}_t$  is used as an input to all classifiers. The proposed method provides the results of the recognition by these classifiers as the final output.

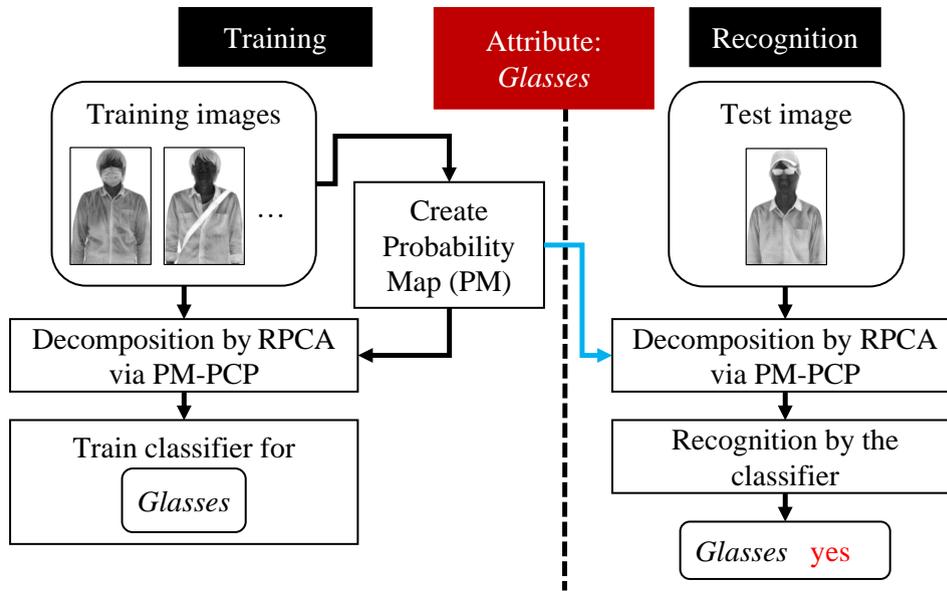


FIGURE 5.3: Process flow of the recognition framework by RPCA via PM-PCP.

### 5.3.2 Recognition Framework with RPCA via PM-PCP

This section describes the recognition framework using RPCA via PM-PCP. The most important modification in this framework compared to the RPCA via PCP introduced in Section 5.3.1, is that the decomposition is performed in a more focused region. The focused region is called the Probability Map. The process flow of the framework is shown in Figure 5.3.

The overall process flow of the framework is divided into training and recognition phases. However, in order to provide a clear description of the framework, the creation process of the Probability Map is explained first. Next, the decomposition process of each phase is described, followed by training/recognition of the classifier.

#### 5.3.2.1 Creation of Probability Map

Every Probability Map created in this framework is unique to each attribute. Positive samples of each attribute are used to create the Probability Maps. These positive samples indicate that the desired attribute is present in the images, otherwise, the images are called negative samples.

First, the positive samples of attribute  $c \in \{c_1, c_2, \dots\}$  are taken from the dataset. The collection of positive samples for attribute  $c$  is represented as  $\mathbf{U}^c = [\mathbf{u}_1^c \mathbf{u}_2^c \dots \mathbf{u}_I^c]$  where  $I$  is the number of positive observations.

For each attribute, one or more rectangular bounding boxes are annotated in the image. For each positive image of attribute  $c$  with a size of  $e \times f$  [pixels], the probability value of each pixel  $o_i^c(e, f) = 1$  for pixels within the rectangular regions and  $o_i^c(e, f) = 0$  otherwise, where  $i = 1, 2, \dots, I$ . After the annotation is performed on all of the positive images, the probability of attribute  $c$  is present in each pixel is calculated by:

$$o^c(e, f) = \frac{1}{I} \sum_{i=1}^I o_i^c(e, f), \quad (5.13)$$

where  $o^c$  for every pixel can be combined into a matrix with a size of  $e \times f$  [pixels], represented by  $\mathbf{O}^c$ . The value of each pixel is the probability for attribute  $c$  calculated in Equation (5.13), ranging from 0 to 1. Although  $\mathbf{O}^c$  can be seen as a probability map, it cannot be used as is for the decomposition. In order to perform the decomposition,  $\mathbf{O}^c$  needs to be rearranged to a column vector  $\mathbf{o}^c$  and duplicated  $n$  times so it has the same dimension as  $\mathbf{M}$ , which is  $m \times n$ . Here,  $m$  represents the number of rows (dimensions), which is  $m = e \times f$ , and  $n$  is the number of observations. This is represented by  $\mathbf{P}^c = [\mathbf{p}_1^c \mathbf{p}_2^c \dots \mathbf{p}_n^c]$  which is the Probability Map that is going to be used for the decomposition, where  $\mathbf{p}_n^c = \mathbf{o}_n^c$ . Figure 5.4 shows examples of Probability Maps of various attributes.

### 5.3.2.2 Decomposition and Recognition Phases

To start the decomposition process, first, images that contain no wearable attribute are categorized in vector form as base data. It is represented as  $\mathbf{B} = [\mathbf{b}_1 \mathbf{b}_2 \dots \mathbf{b}_H]$  with  $H$  observations. The base data is used both in the training and the recognition phases.

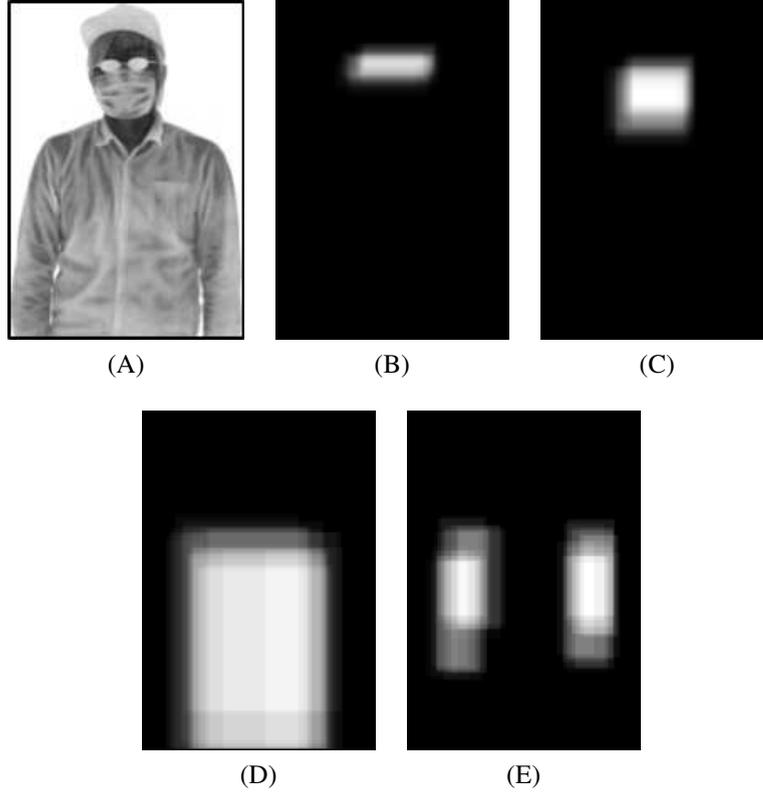


FIGURE 5.4: Example of thermal image and various Probability Maps (PMs). Each image represents: (A) Thermal image with *hat*, *glasses*, and *mask* attributes; (B) *Glasses* PM; (C) *Mask* PM; (D) *Shoulder bag* PM; (E) *Backpack* PM

The training data of attribute  $c$  is represented as  $\mathbf{W}^c = [\mathbf{w}_1^c \mathbf{w}_2^c \cdots \mathbf{w}_j^c]$  with  $J$  observations. This training data can be considered as the positive samples  $\mathbf{U}^c$  with the addition of the negative samples, where  $I \leq J$ . Therefore, for every attribute  $c$ , the attribute-specific training data  $\mathbf{W}^c$  will then be decomposed using PM-PCP as follows:

$$\begin{bmatrix} \mathbf{B} & \mathbf{w}_j^c \end{bmatrix} = \mathbf{M}_j = \mathbf{L}'_j + (\mathbf{P}^c \circ \mathbf{S})_j, \quad (5.14)$$

where

$$(\mathbf{P}^c \circ \mathbf{S})_j = \begin{bmatrix} \mathbf{s}_1^c & \mathbf{s}_2^c & \cdots & \mathbf{s}_H^c & \mathbf{x}_j^c \end{bmatrix}. \quad (5.15)$$

The entry  $\mathbf{x}_j^c$  in Equation (5.15) is the sparse representation that corresponds to  $w_j^c$ . The decomposition process of training data is performed  $H$  times, and the sparse results are grouped together as  $\mathbf{X}^c = [\mathbf{x}_1^c \ \mathbf{x}_2^c \ \cdots \ \mathbf{x}_H^c]$  for the training of the classifier.

Since the training data  $\mathbf{X}^c = [\mathbf{x}_1^c \ \mathbf{x}_2^c \ \cdots \ \mathbf{x}_H^c]$  contains positive and negative samples of attribute  $c$ , any method can be used as the classifier in this recognition framework. Note that one classifier is trained for each attribute. In the case of a binary classifier, both the positive and negative samples of attribute  $c$  are provided until it is capable to classify them. When the classifier is trained, it marks the end of the training phase.

For the recognition phase, the proposed method needs to recognize attribute  $c$  in an input thermal image  $t$ . PM-PCP is then used to solve the following:

$$\begin{bmatrix} \mathbf{B} & t \end{bmatrix} = \mathbf{M}_t = \mathbf{L}'_t + (\mathbf{P}^c \circ \mathbf{S})_t, \quad (5.16)$$

where

$$(\mathbf{P}^c \circ \mathbf{S})_t = \begin{bmatrix} s_1^c & s_2^c & \cdots & s_H^c & \mathbf{x}_t^c \end{bmatrix}. \quad (5.17)$$

Note that Equation (5.17) includes an entry  $\mathbf{x}_t^c$ , which is the sparse representation of  $t$  for attribute  $c$ . This entry is used as the input for the classifier, which was trained for attribute  $c$ . The classifier then will determine whether the attribute  $c$  is present in  $t$  or not. In the case of a binary classifier, the output will be 1 when the attribute  $c$  is present, and 0, otherwise.

## 5.4 Experiments and Analysis

To assess the decomposition capability of the proposed method, experiments were conducted by evaluating the decomposed images and the recognition results. This section starts by introducing the dataset used for the experiment, followed by the

setup of the experiment elaborated in the following subsection. The next subsection describes some other methods as comparison. Finally, the decomposition and recognition results and the analysis are provided in the last subsection.

### 5.4.1 Dataset

Since a public thermal infrared image dataset that contains various attributes was not readily available, a private dataset was newly created for this research. It consists of upper body images from 14 people (11 males and three females in the age range of 20–31), with up to seven different wearable attributes per person. Subjects were standing at a distance of 3.0 to 3.5 m from the camera, and only the upper region of the body was captured by the thermal infrared camera. The attributes available in the dataset and the number of their positive images available can be seen in Table 5.1.

A total of 408 images were captured in the thermal infrared spectrum. The images were taken indoors at room temperature (around 22–25 degrees Celsius), using an Avionics' TVS-500EX camera [113] again, which can capture infrared wavelengths from 8 to 14  $\mu\text{m}$ . The camera was set to capture infrared radiation whose temperature ranges from 25 to 36 degrees Celsius. The camera is able to take pictures from both the thermal infrared and the visible spectra simultaneously, although the focal axes are slightly different. However, only the thermal infrared images were used in this experiment. The raw image size was  $320 \times 240$  pixels for the thermal infrared spectrum.

Various preprocessings were performed to these thermal infrared images. First, the human body region in each image is manually cropped. Specifically for the creation of the Probability Map, the regions where the attributes are located were manually annotated. Then, the size of these upper body images were averaged to  $140 \times 204$  pixels. Note that images in the dataset are shown in the “hotblack” color scheme, which means the closer the pixel is to 0 (black), the hotter the temperature is. Examples of images from the dataset are shown in Figure 5.5.

TABLE 5.1: Distribution of the seven wearable attributes in the dataset. Note that there are images that contain more than one wearable attributes.

Attributes	# of images
No attribute	28
<i>Glasses</i>	168
<i>Mask</i>	168
<i>Hat</i>	112
<i>Helmet</i>	80
<i>Hoodie</i>	40
<i>Shoulder bag</i>	48
<i>Backpack</i>	40
Total	694

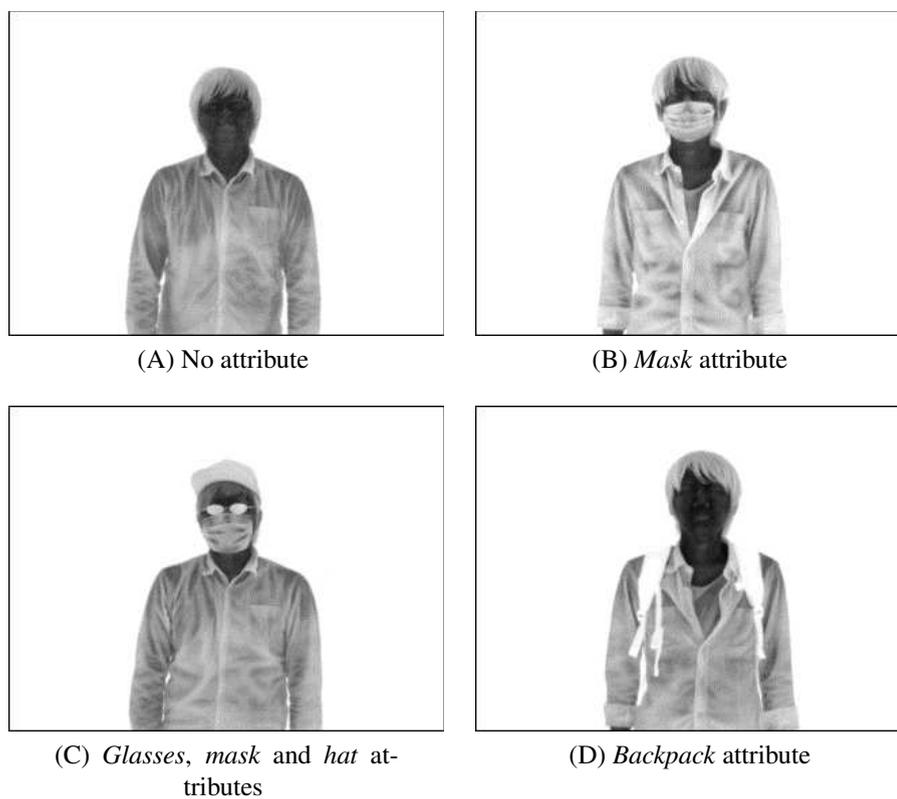


FIGURE 5.5: Image examples from the dataset.

### 5.4.2 Experimental Setup

In this experiment, the classifier of choice was Support Vector Machine (SVM) [92], with the decomposition performed on a pixel value. This means that no feature extraction was performed on the image. The way the data was divided ensures the training and test data do not intersect. The experiment was performed with a leave-one-person-out cross-validation.

To evaluate the capability of the proposed methods, F-score was used as measures. The decomposition results were used as inputs to the SVM classifier, and the performance for each decomposition method was analyzed. Let us assume that we wish to recognize the *glasses* attribute. In this case, the positive samples are images where *glasses* is present and the negative samples are images where *glasses* is not present. When the SVM successfully recognizes *glasses* from a positive sample (output 1), it is a True Positive case. On the other hand, a True Negative case occurs when the SVM recognizes that there is no *glasses* from a negative sample (output 0). Based on this binary output, precision, recall, and consequently the F-score can be calculated. We can then compare the F-score of each decomposition method for attribute recognition.

### 5.4.3 Comparative Methods

The experiment compares the capability of RPCA via PCP and RPCA via PM-PCP for decomposition with other methods. It was conducted in the same way for all of the methods to guarantee fairness. The comparative methods used in this experiment are provided in the following list. The summary of all comparative methods is shown in Table 5.2.

- **No Decomposition:** No decomposition is performed in this method, as the name implies. Thermal images are directly used for training and testing of the classifier. This method serves as the baseline.

TABLE 5.2: Comparison of the methods performed in the experiment.

Method	PM	Decomposition Method
No Decomposition	—	Not Available
Conventional Average	—	Subtraction between $\bar{\mathbf{b}}$ and $\mathbf{t}$
Conventional PCA	—	Subtraction between $\tilde{\mathbf{t}}$ and $\mathbf{t}$
RPCA via PCP (Proposed 1)	—	PCP
Post-PCP PM (Proposed 2)	✓	PCP first, then apply PM
RPCA via PM-PCP (Proposed 3)	✓	PM-PCP

- **Conventional Average:** The decomposition is performed utilizing the base data  $\mathbf{B}$ . First, the average vector is calculated from the base data  $\mathbf{B}$ , represented as  $\bar{\mathbf{b}}$ . Then, the “decomposition” is performed by subtracting  $\bar{\mathbf{b}}$  from the test data  $\mathbf{t}$ .
- **Conventional PCA:** The decomposition is performed utilizing Principal Component Analysis (PCA). First, the projection matrix  $\mathbf{V}$  is calculated from the base data  $\mathbf{B}$  by PCA. Next, the test data  $\mathbf{t}$  is projected to and back from the Eigenspace, represented by  $\tilde{\mathbf{t}}$ . Then, the “decomposition” is performed by subtracting  $\tilde{\mathbf{t}}$  from the test data  $\mathbf{t}$ .
- **Post-PCP PM (Proposed 2):** The decomposition is performed utilizing the Probability Map. First, the decomposition is performed just like the RPCA via PCP method. Then, the Probability Map is applied on  $\mathbf{x}_t$ , which is the sparse entry that corresponds to the test data  $\mathbf{t}$ . This is a simplified version of using the Probability Map without modifying the RPCA via PCP problem.

TABLE 5.3: Comparison of the proposed methods with other methods in F-Score.

Methods	Attributes									Average over attributes in:		
	<i>Glasses</i>	<i>Mask</i>	<i>Hat</i>	<i>Helmet</i>	<i>Shoulder bag</i>	<i>Backpack</i>	<i>Head region</i>	<i>Torso region</i>	<i>All regions</i>			
No Decomposition	0.607	0.603	0.471	0.357	0.222	0.200	0.510	0.211	0.410			
Conventional Average	0.834	0.988	0.551	0.800	0.538	0.849	0.793	0.694	0.760			
Conventional PCA	0.854	0.985	0.476	0.645	0.698	0.892	0.740	0.795	0.758			
RPCA via PCP (Proposed 1)	0.861	0.982	0.502	0.754	<b>0.723</b>	<b>0.947</b>	0.775	<b>0.835</b>	0.795			
Post-PCP PM (Proposed 2)	0.918	<b>0.997</b>	0.581	0.843	0.675	0.841	0.835	0.758	0.809			
RPCA via PM-PCP (Proposed 3)	<b>0.919</b>	0.991	<b>0.623</b>	<b>0.868</b>	0.692	0.868	<b>0.850</b>	0.780	<b>0.827</b>			

#### 5.4.4 Results and Analysis

Table 5.3 shows the recognition results of all methods, including results of each attribute and average over attributes on the head region, torso region, and as a whole. Bar graph representations are also shown in Figures 5.6 and 5.7 for results of each attribute and each region, respectively. Based on these results, a conclusion can be made that the approach of performing decomposition on the thermal infrared images showed a significant performance increase.

We can see that the results of both Conventional Average and Conventional PCA methods showed improvement compared to the baseline, but the RPCA via PCP method outperformed both methods in the case of decomposition without prior knowledge. This shows that the decomposition idea in general works well.

Overall, the decomposition by RPCA via PM-PCP achieved the highest performance with an average of 0.827 F-score, outperforming all the other methods. Coming in second was the Post-PCP PM method, which confirmed the usefulness of the PM for the decomposition. Other than the *mask* attribute, it failed to outperform the RPCA via PM-PCP method.

However, the RPCA via PM-PCP method was not the best for attributes in the torso region. Both RPCA via PCP and Conventional PCA methods performed better in this case. Except for the *backpack* attribute where the RPCA via PCP method held a clear advantage, the differences between the three methods were relatively small.

The reason is most likely due to the variations of the shape and the height of the human bodies in the dataset which in turn made the rectangular region for the attribute labeling not as accurate. The by-product of this is that since the probability of some parts in the *backpack* Probability Map were relatively low, the weighted decomposition made the sparse representation of the attribute information of those parts not as apparent when compared to those of the Post-PCP PM method. This can be observed in the bottom row of Figure 5.8, where the qualitative comparison is shown.

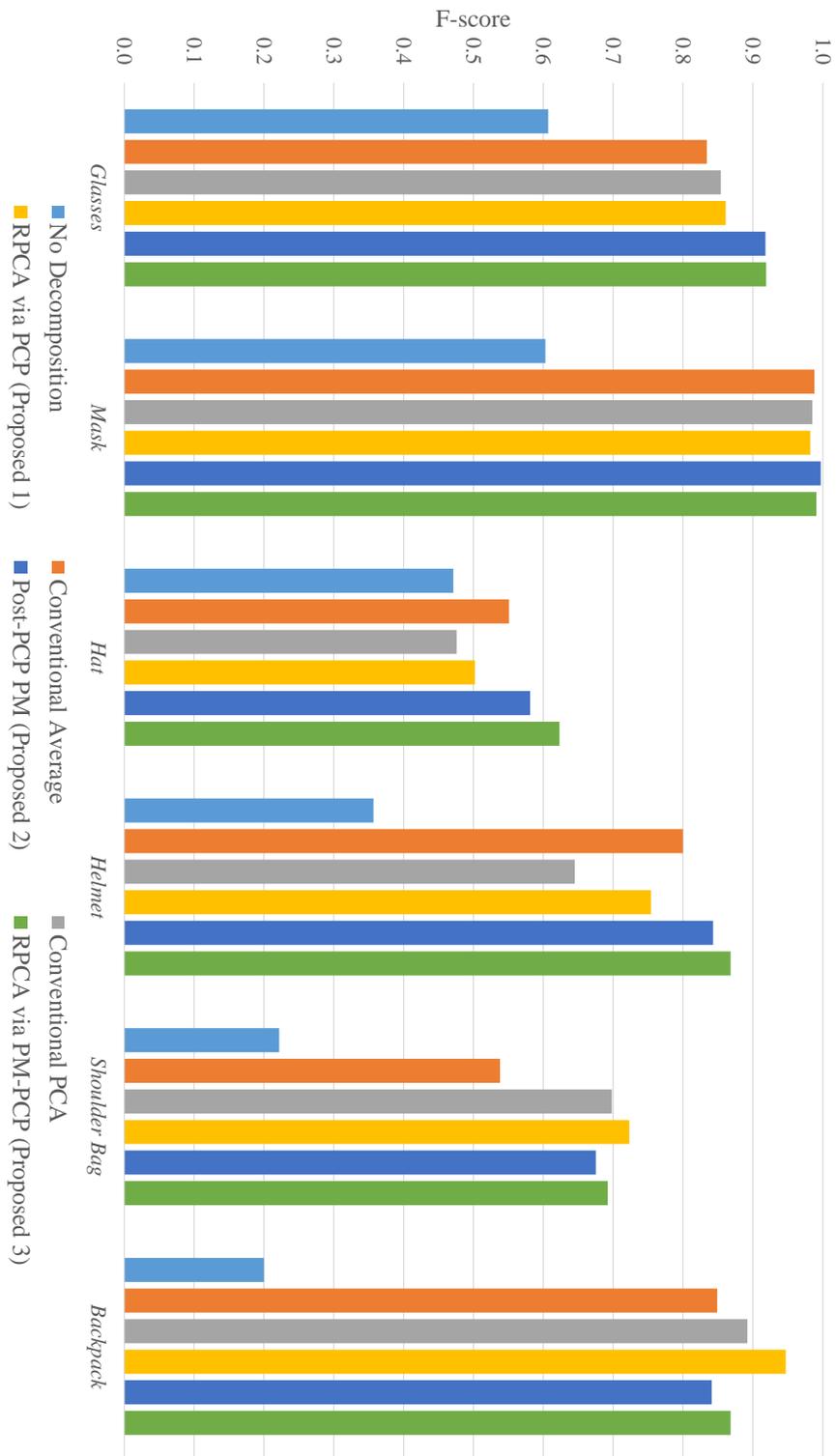


Figure 5.6: Evaluation results by F-score per attribute.

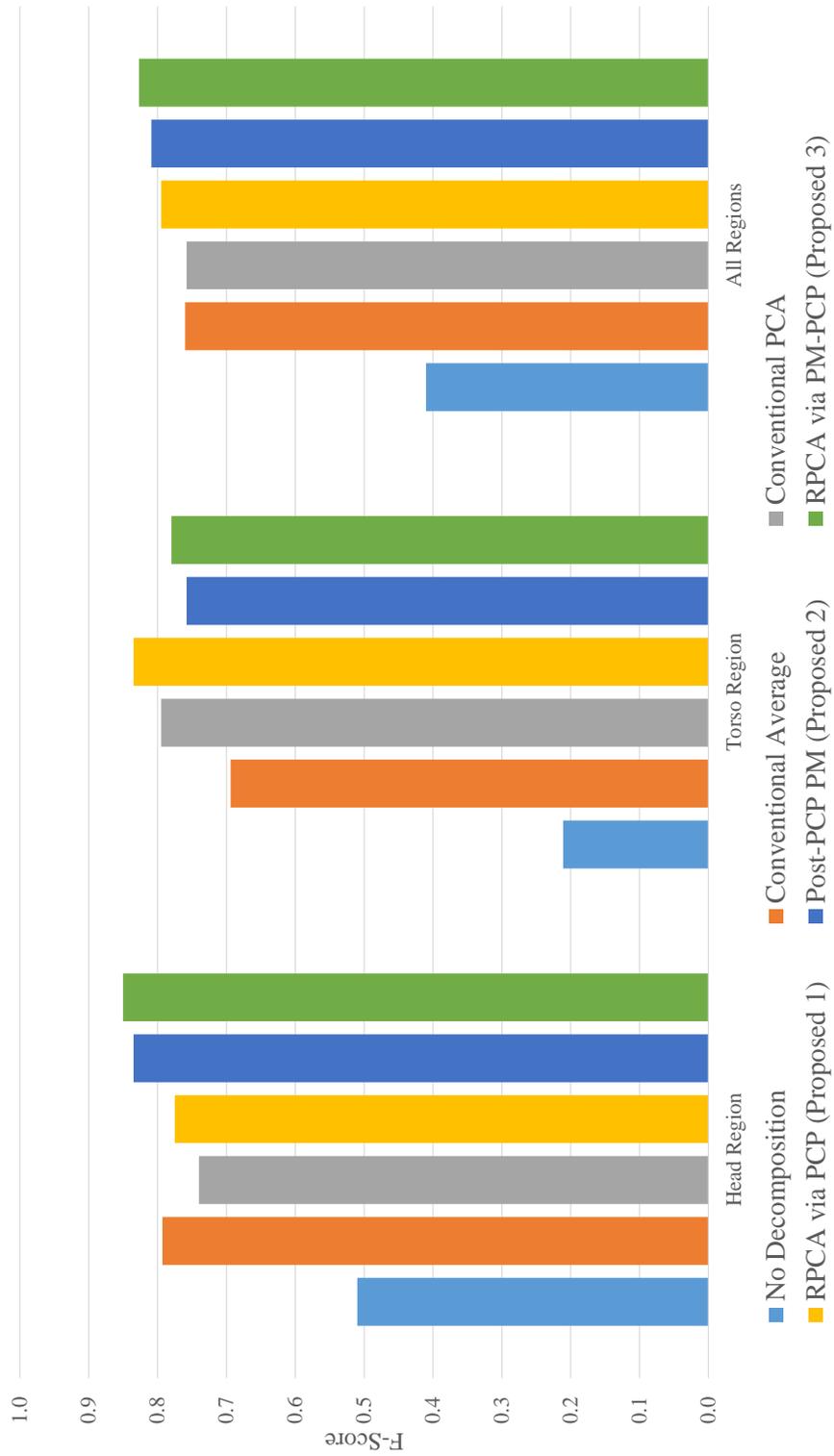


FIGURE 5.7: Evaluation results by F-score per region.

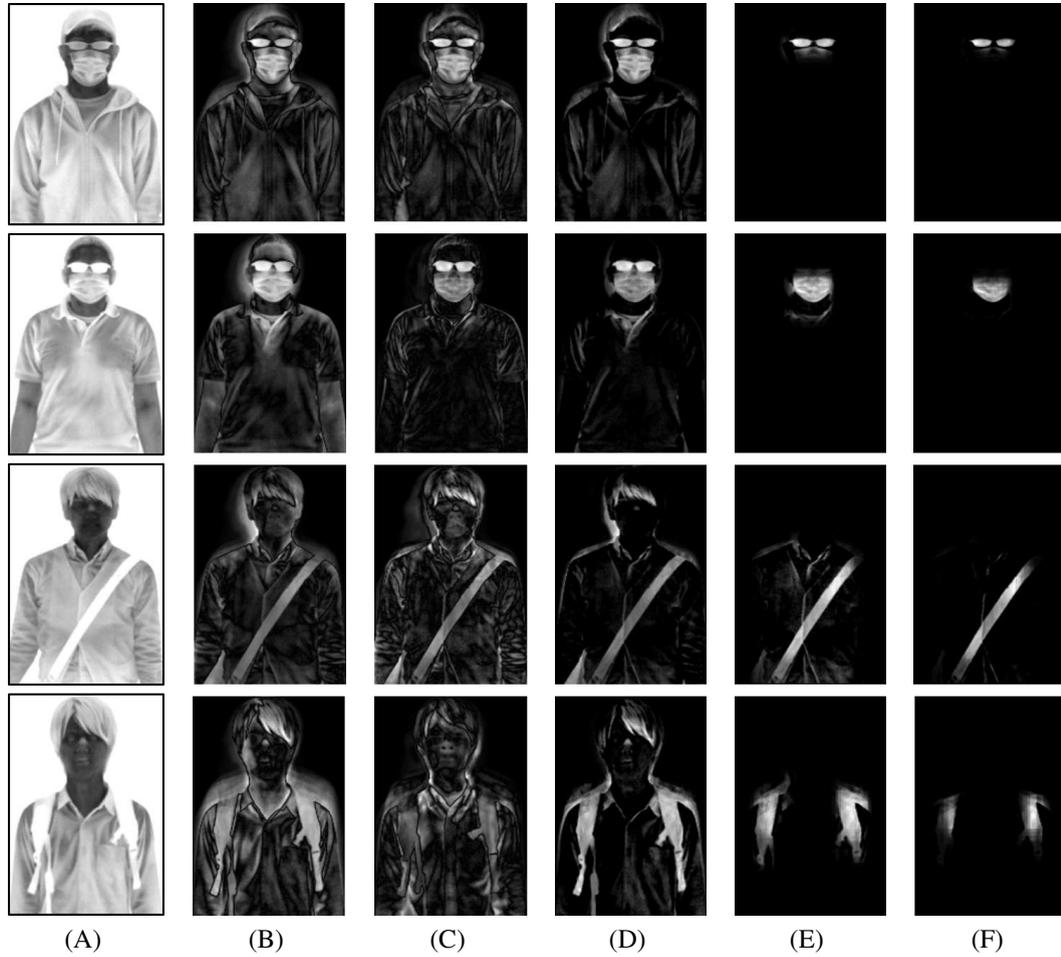


FIGURE 5.8: Examples of decomposition results by various methods. Images represent: (A) Original thermal infrared; (B) Results of Conventional Average; (C) Results of Conventional PCA; (D) Results of RPCA via PCP; (E) Results of Post-PCP PM; (F) Results of RPCA via PM-PCP. The decomposition is made for (from top to bottom): *glasses*, *mask*, *shoulder bag*, and *backpack* attributes. For visualization purpose, the minimum and maximum pixel values for each image were normalized to 0 and 1, respectively.

Note that the images in Figure 5.8 have been normalized to  $[0, 1]$  for visualization purposes.

In Figure 5.8, the sparse representation of the thermal infrared image clearly shows the extracted attributes decomposed by each method. The Conventional Average method produced a lot more noise in the results than other decomposition methods, where the body shape can still be clearly seen. The Conventional PCA method

showed better decomposition results than the Conventional Average method, although not by much. The results of the RPCA via PCP method were better than the previous methods, although some clothes and hair details were still noticeable.

The results of the Post-PCP PM method showed a more focused decomposition. However, there is still noise unrelated to the attribute the proposed method is trying to recognize. The decomposition by the RPCA via PM-PCP method successfully extracted the attribute information almost exclusively. The usage of Probability Map proved to be helpful for the decomposition, as shown in both the quantitative and qualitative results. Although the results of the RPCA via PM-PCP method had some erosion on the attributes, the Post-PCP PM method contained more non-attribute noise, which is detrimental to the recognition performance.

For further discussion, results of decomposition by methods that use Probability Map in the case of negative sample used as the test image are provided. These are the cases where the decomposed image does not contain the attribute the proposed method is trying to recognize. For example, take the case where the input image contains only the *backpack* attribute when the decomposition was performed for the *glasses* attribute. Figure 5.9 shows some of the results of such decompositions. Note that once again, the decomposition results were normalized to  $[0, 1]$  for visualization purposes. It can be observed that there are details of clothing and face in the decomposition results of both methods. However, the decomposition by the RPCA via PM-PCP method produced better results with less noise than the Post-PCP PM method. These noise are clothes and face details, which is not necessary for attribute recognition. This leads to lower recognition results when compared to the results of the RPCA via PM-PCP method.

## 5.5 Summary

This chapter addressed the answer to the second research question in this thesis, which was “What is the person wearing?” The chapter tackled a wearable attribute

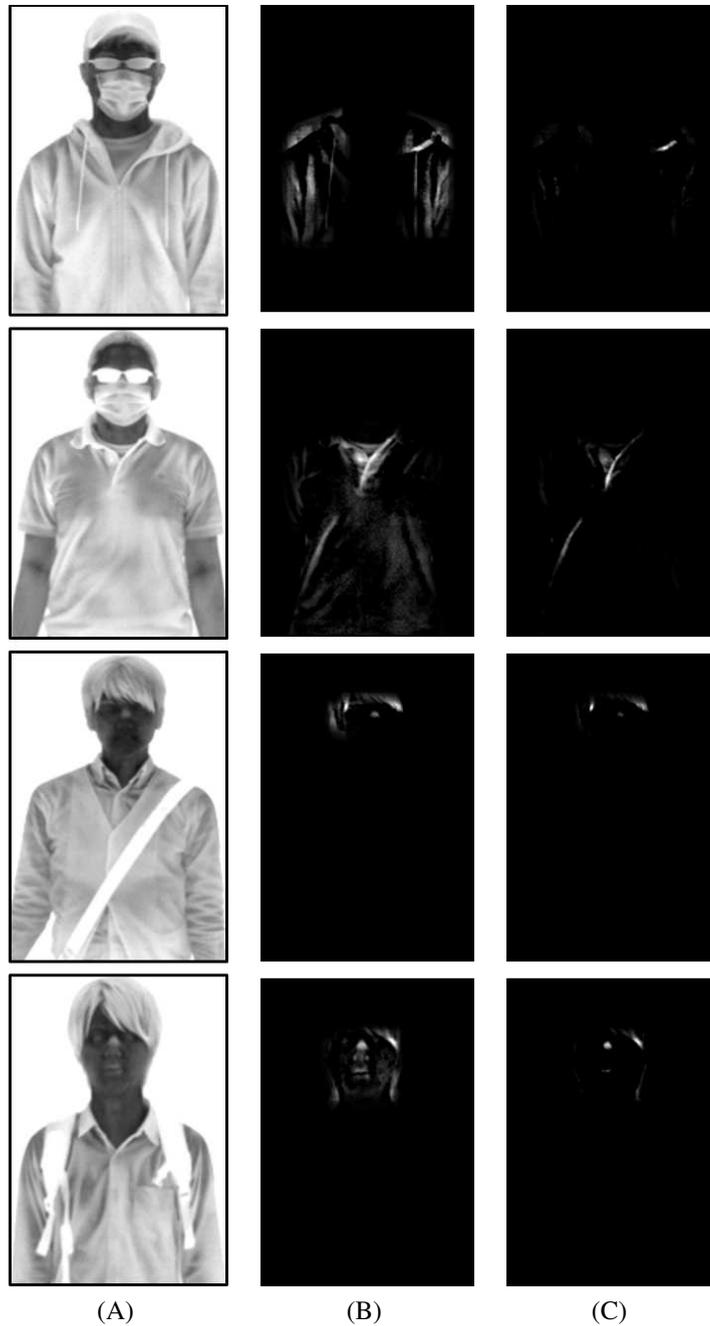


FIGURE 5.9: Examples of decomposition results by methods that use Probability Map in the case of negative input images. Images represent: (A) Original thermal infrared; (B) Results of Post-PCP PM; (C) Results of RPCA via PM-PCP. The decomposition is made for (from top to bottom): *backpack*, *shoulder bag*, *glasses*, and *mask* attributes. For visualization purpose, the minimum and maximum pixel values for each image were normalized to 0 and 1, respectively. Since the images do not contain the target attributes, pixels of the decomposed images are close to black.

---

recognition task in thermal infrared images to answer this question. In order to achieve this, two versions of decomposition-based methods using RPCA were proposed; The Robust PCA via PCP method and its modification, the Robust PCA via PM-PCP method. Probability Map was introduced as prior knowledge to achieve a more focused decomposition for the second proposed method.

Experiments were performed to evaluate the attribute recognition performance with the decomposition approach. The results showed that the decomposition by Robust PCA via PM-PCP achieved the highest performance in average with a 0.827 F-score, outperforming all other methods. By visual observation, the resulting images also showed relatively better extracted attributes.



# Chapter 6

## Conclusion

This chapter provides the conclusion of the thesis. The overall summary is shown in Section 6.1. Section 6.2 describes additional works that can be performed in the future. Section 6.3 provides some closing remarks at the end to finish the thesis.

### 6.1 Thesis Summary

This thesis attempted to gain insight and understanding to human images in the thermal infrared spectrum. Since thermal infrared spectrum is different from the visible spectrum, it is difficult for humans to observe and understand the environment in this spectrum. This also holds true in the case of observing human images captured in the thermal infrared spectrum, because features available in the visible spectrum are not necessarily available in this spectrum. This shows our lack of understanding of thermal infrared images. However, it is not a sufficient reason to ignore the advantages of using these images. Therefore, instead of trying to force humans to increase their understanding of thermal infrared images, the task of understanding can be given to a system. If a system can understand the thermal infrared images and feed us useful information from them, to a certain extent, we can forsake understanding the images

ourselves. To achieve this, two research topics were tackled in this thesis to gain understanding in both visual and semantic senses from thermal infrared images.

The first research topic was the visual understanding, which was performed by reconstructing face images from the thermal infrared to the visible spectra. This reconstruction is very beneficial in the case of using a thermal infrared camera for surveillance, because humans struggle to recognize faces taken in this spectrum. When an automated surveillance system recognizes a person as a wanted individual, affirmation of this conjecture is necessary. By reconstructing the face to the visible spectrum, humans can confirm this personally and decide the next action. Since the biometric information of a face is unique to a person regardless of the spectrum it is in, the system can infer the correlation of a face between different spectra. The reconstruction can then be performed based on this relationship between these face images taken in both the thermal infrared and the visible spectra.

The reconstruction method proposed in this topic was composed of two steps. In the first step, it reconstructs a holistic face image using a low-rank representation. The second step includes reconstruction of patches of the face image to reintroduce details and characteristics lost in the first step. Experiments were conducted to evaluate the reconstruction performance. The proposed method outperformed other comparative methods and achieved 33.11 dB, 0.95 and 98.44% in PSNR, SSIM, and face recognition evaluations, respectively.

The second research topic was the semantic understanding of the human image taken in the thermal infrared spectrum. Although there are many semantic meanings that can be obtained from a human image, this research specifically wanted to recognize the existence of wearable attributes in the human image. In the case where human identification is automated, wearable attributes on the face of a person occlude the identification process, which means that the identity of the person becomes unknown. In areas where the identity of a person is important, such as banks and military outposts, this unknown person could be a security risk. Ideally, it is better for such a person to not enter these areas, in which case a preventative measure is

useful. Therefore, identifying what attributes people are wearing before they enter is important. Since only the information of the attributes is necessary from an image in order to achieve this, a decomposition approach for recognition was proposed. The decomposition approach obtained only the attribute information by taking advantage of the properties of the wearable attributes in the thermal infrared spectrum and used them for the attribute recognition process.

The decomposition was achieved by the innovative use of Robust Principal Component Analysis (RPCA), where the proposed method exercised control over the input data. To improve the proposed method, prior knowledge was introduced in the form of Probability Map. This prior knowledge shows the likelihood of an attribute being present in a certain region to achieve a more focused decomposition. Experiments were conducted to evaluate the recognition performance with the decomposition approach. The proposed PM-PCP method outperformed other comparative methods with a 0.827 F-score in average.

Based on the two aforementioned research topics tackled in this thesis, two systems that can provide insights to face and human upper body images in the thermal infrared spectrum were implemented. Face image reconstruction of the first research topic provides visual understanding of what a face image taken in the thermal infrared spectrum looks like in the visible spectrum. Wearable attribute recognition of the second research topic provides semantic understanding to human upper body images, providing humans information of various objects worn by the person in an image. There are still many areas to improve upon, but these systems can be useful in its application for security surveillance system.

## **6.2 Future Work**

Although some understanding of human images in the thermal infrared spectrum have been made, there are still researches that can be conducted to gain both visual and semantic understandings. This means that more questions can be asked in

addition to the two research questions mentioned in Chapter 1.

From visual appearance, the immediate information that can be obtained is temperature. Normal body temperature of an adult human is approximately 37 degrees Celsius, ranging from 36.2 to 37.5 degrees Celsius [134]. Although the work by Wunderlich [134] was published in 1871, this standard has been widely accepted. However, the human body is not that simple. There are many factors that can affect the body temperature of a human [135, 136]. Additionally, medical technology has also advanced, hence the ranges of normal body temperature need to be adjusted [137]. Nevertheless, anyone with temperature outside of a normal one might make us ask a question, “Is the person sick or healthy?” We can already see an application of this in airports to find people with fever. Automation and further research on this topic warrants more practical use.

There are also other semantic understandings that can be obtained from images. For example, questions can be made about other attributes in a human image such as “Is the person’s hair long, short, or bald?” and “What kind of clothing is the person wearing?” These questions need a system to take a closer look at the shapes and understand them before answering is possible.

As for the two research topics tackled in this thesis, their possible future works are provided in the following description.

**Topic 1:** Currently, the visual understanding of faces known by the system produces great results. However, in reconstructing the face of an unknown person, the proposed method faces difficulty. In the future, it is important to consider additional improvements in this case. It might be possible to do this by learning the relationship between the visible and the thermal infrared spectra non-linearly using a kernel such as Kernel Canonical Correlation Analysis (KCCA) [117–121], since currently the proposed method assumes the relationship to be linear. Another area of improvement is to consider the usage of weighting when combining the patches and the globally reconstructed images.

**Topic 2:** In the current situation, the results of the decomposition method based on decomposition are relatively good. For future work, the level of difficulty of the research can be increased. For example, the usage of non-frontal images to add more challenges. In a real surveillance situation, having consistent frontal images are relatively difficult due to camera placements and other human factors. The difficulty of utilizing different angle and pose information is well-known and some modifications need to be adapted in order to reach satisfying performance. This fact warrants further research on this topic.

The challenges of utilizing different angle and pose information also applies to the first research topic. The ideal capture condition rarely occurs in an unconstrained environment, hence further experimentations and improvements need to be conducted to handle the difficulties in such environment.

Lastly, both research topics made an assumption that a detection process took place beforehand, and take only face or upper body images as their input. Since the methods proposed in both research topics are relatively sensitive to image translations, the detector accuracy may also affect the overall performance of proposed methods. Therefore, the detection process itself can also be an interesting work to focus on.

## **6.3 Closing Remarks**

As closing words, it needs to be reiterated that there are many researches done in the visible spectrum, and there are not enough of them in the thermal infrared spectrum in comparison. Since the information available in the thermal infrared spectrum is different than those in the visible spectrum, there are many opportunities to expand our understanding of thermal infrared images. Although this thesis may only be a small drip in the ocean of knowledge, hopefully it can incite discussions and more researches in the thermal infrared spectrum.

One of the big aspirations of this thesis is to have thermal infrared cameras as a complement or more, as a substitute to normal cameras. There are many factors

that need to be considered for that purpose. The first and the foremost is the cost of thermal infrared cameras. While the trend of the price keeps going down, the costs of these cameras in general are higher than normal cameras. Additionally, some of the high-end cameras that have high image resolutions are even more expensive. Having said that, normal cameras chosen for surveillance are usually not the high-end models. Therefore, the usage of low- to mid-end thermal infrared cameras can also be considered for surveillance.

In regards to the thermal infrared image, one of the most important factor to consider is the lack of an international standard. Since temperature is an integral part of imaging process in this spectrum, a standard on observed range of temperature for an image is important. In the current situation, a lack of consensus on the observed range of temperature causes other researches in the thermal infrared spectrum a challenge to recreate. A publicly available dataset can be used to avoid this problem, but there are only few of them. In addition, the datasets that are available may not be easily adaptable to various researches due to their settings.

Although there may be more factors to consider, the last factor considered in this research is the familiarity of people to the thermal infrared spectrum. In other words, the exposure of thermal infrared spectrum to the masses. With sufficient familiarity and exposure, people will be more likely to buy thermal cameras, increasing the demand. In turn, supply may raise and decrease the overall price of the cameras. In addition, people may be more interested to conduct researches in this spectrum.

In the end, having both of normal and thermal infrared cameras is advantageous in regards to surveillance since monitoring in multiple spectra can cover the weaknesses of each spectrum compared to when using only one of them. Therefore, although substituting visible cameras is a massive challenge, popularizing thermal infrared as a complement is still viable in the future.

# Bibliography

- [1] R. Feris, R. Bobbitt, L. Brown, and S. Pankanti. Attribute based people search: Lessons learnt from a practical surveillance system. In *Proceedings of the 4th ACM International Conference on Multimedia Retrieval*, pages 153–160, April 2014.
- [2] J. W. Davis and V. Sharma. Robust detection of people in thermal imagery. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 4, pages 713–716, August 2004.
- [3] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages I-11–I-18, June 2003.
- [4] K.-C. Lee, J. Ho, and D. J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, May 2005.
- [5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, February 2009.
- [6] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister. Statistics of infrared images. In *Proceedings of 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–7, June 2007.

- [7] N. Nandhakumar and J. K. Aggarwal. Integrated analysis of thermal and visual images for scene interpretation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):469–481, July 1988.
- [8] W. K. Wong, P. N. Tan, C. K. Loo, and W. S. Lim. An effective surveillance system using thermal camera. In *Proceedings of 2009 International Conference on Signal Acquisition and Processing*, pages 13–17, April 2009.
- [9] W. K. Wong, Z. Y. Chew, C. K. Loo, and W. S. Lim. An effective trespasser detection system using thermal camera. In *Proceedings of the 2nd International Conference on Computer Research and Development*, pages 702–706, May 2010.
- [10] M. Shah, O. Javed, and K. Shafique. Automated visual surveillance in realistic scenarios. *IEEE MultiMedia*, 14(1):30–39, January 2007.
- [11] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani. Computer vision system for in-house video surveillance. *IEE Proceedings — Vision, Image and Signal Processing*, 152:242–249, April 2005.
- [12] A. R. Dick and M. J. Brooks. Issues in automated visual surveillance. In *Proceedings of the 7th International Conference on Digital Image Computing: Techniques and Applications*, volume 1, pages 195–203, December 2003.
- [13] H. M. Dee and S. A. Velastin. How close are we to solving the problem of automated visual surveillance? *Machine Vision and Applications*, 19(5–6):329–343, October 2008.
- [14] D. A. Vaquero, R. S. Varis, D. Tran, L. Brown, A. Hampapur, and M. Turk. Attribute based people search in surveillance environment. In *Proceedings of IEEE Computer Society Workshop on Applications of Computer Vision 2009*, pages 1–8, December 2009.
- [15] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, September 1933.

- [16] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4): 321–377, December 1936.
- [17] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.
- [18] E. J. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11.1–11.37, May 2011.
- [19] W. M. Haynes. *CRC Handbook of Chemistry and Physics, 9th Edition*. CRC Handbook of Chemistry and Physics. Taylor & Francis, June 2011.
- [20] E. Rutherford. *Radioactive Substances and Their Radiations*. University Press, 1913.
- [21] W. C. Röntgen. On a new kind of rays. *Science*, 3(59):227–231, February 1896.
- [22] G. Laufer. *Introduction to Optics and Lasers in Engineering*. Cambridge University Press, July 1996.
- [23] H. Bradt. *Astronomy Methods: A Physical Approach to Astronomical Observations*. Cambridge Planetary Science. Cambridge University Press, December 2003.
- [24] L. Ohannesian and A. Streeter. *Handbook of Pharmaceutical Analysis*. Drugs and the Pharmaceutical Sciences. Taylor & Francis, November 2001.
- [25] M. C. Dash and S. P. Dash. *Fundamentals of Ecology 3rd Edition*. Tata McGraw-Hill Education, June 2009.
- [26] R. T. Hitchcock and American Industrial Hygiene Association. Non-Ionizing Radiation Committee. *Radio-frequency and Microwave Radiation*. Nonionizing radiation guide series. American Industrial Hygiene Association, June 2004.

- [27] G. A. Jones, D. H. Layer, T. G. Osenkowsky, and E. A. Williams. *National Association of Broadcasters Engineering Handbook: NAB Engineering Handbook 10th Edition*. Taylor & Francis, April 2013.
- [28] ISO/TC 172 Optics and Photonics. *ISO 20473 Spectral Bands*. International Organization for Standardization (ISO), April 2007.
- [29] M. Rowan-Robinson. *Night Vision: Exploring the Infrared Universe*. Cambridge University Press, April 2013.
- [30] A. D'Amico, C. D. Natale, F. L. Castro, S. Iarossi, A. Catini, and E. Martinelli. Volatile compounds detection by IR acousto-optic detectors. In *2008 NATO Advanced Study Institute on Unexploded Ordnance Detection and Mitigation*, pages 21–59, July–August 2008.
- [31] D. H. Sliney, R. T. Wangemann, J. K. Franks, and M. L. Wolbarsht. Visual sensitivity of the eye to infrared laser radiation. *Journal of the Optical Society of America*, 66(4):339–341, April 1976.
- [32] M. Planck. *The Theory of Heat Radiation*. P. Blakiston's Son & Co, February 1914.
- [33] M. Massoud. *Engineering Thermofluids: Thermodynamics, Fluid Mechanics, and Heat Transfer*. Springer Berlin Heidelberg, December 2005.
- [34] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska. A new face database simultaneously acquired in visible, near-infrared and thermal spectrums. *Cognitive Computation*, 5(1):119–135, March 2013.
- [35] F. J. Prokoski, R. B. Riedel, and J. S. Coffin. Identification of individuals by means of facial thermography. In *Proceedings of 1992 International Carnahan Conference on Security Technology: Crime Countermeasures*, pages 120–125, October 1992.

- [36] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi. Recent advances in visual and infrared face recognition — A review. *Computer Vision and Image Understanding*, 97(1):103–135, June 2005.
- [37] M. Akhloufi, A. Bendada, and J.-C. Batsale. State of the art in infrared face recognition. *Quantitative InfraRed Thermography Journal*, 5(1):3–26, January 2008.
- [38] R. S. Ghiass, A. Bendada, and X. Maldague. Infrared face recognition: A review of the state of the art. In *Proceedings of the 10th International Conference on Quantitative InfraRed Thermography*, pages 47.1–47.8, July 2010.
- [39] C. San Martin, R. Carrillo, P. Meza, H. Mendez-Vazquez, Y. Plasencia, E. García-Reyes, and G. Hermosilla. Recent advances on face recognition using thermal infrared images. In *Reviews, Refinements and New Ideas in Face Recognition*, pages 95–112. InTech, July 2011.
- [40] R. S. Ghiass, O. Arandjelović, H. Bendada, and X. Maldague. Infrared face recognition: A literature review. In *Proceedings of 2013 International Joint Conference on Neural Networks*, pages 1–10, August 2013.
- [41] D. A. Socolinsky, L. B. Wolff, J. D. Neuheisel, and C. K. Eveland. Illumination invariant face recognition using thermal infrared imagery. In *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages I-527–I-534, December 2001.
- [42] S. Wang, Z. Liu, P. Shen, and Q. Ji. Eye localization from thermal infrared images. *Pattern Recognition*, 46(10):2613–2621, October 2013.
- [43] B. Martinez, X. Binefa, and M. Pantic. Facial component detection in thermal imagery. In *Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition*, pages 48–54, June 2010.

- [44] T. Ojala, M. Pietikainen, and D. Hardwood. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, volume 1, pages 582–585, October 1994.
- [45] D. Bhattacharjee, A. Seal, S. Ganguly, M. Nasipuri, and D. K. Basu. A comparative study of human thermal face recognition based on Haar wavelet transform and local binary pattern. *Computational Intelligence and Neuroscience*, 2012:261089.1–261089.12, June 2012.
- [46] P. Buddharaju and I. Pavlidis. Physiological face recognition is coming of age. In *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 128–135, June 2009.
- [47] R. S. Ghiass, O. Arandjelovic, H. Bendaday, and X. Maldague. Vesselness features and the inverse compositional AAM for robust face recognition using thermal IR. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pages 357–364, July 2013.
- [48] A. Seal, D. Bhattacharjee, M. Nasipuri, and D. K. Basu. Minutiae based thermal face recognition using blood perfusion data. In *Proceedings of 2011 International Conference on Image Information Processing*, pages 1–4, November 2011.
- [49] A. Seal, S. Ganguly, D. Bhattacharjee, M. Nasipuri, and D. K. Basu. Automated thermal face recognition based on minutiae extraction. *International Journal of Computational Intelligence Studies*, 2(2):133–156, July 2013.
- [50] T. Bourlai, A. Ross, C. Chen, and L. Hornak. A study on using mid-wave infrared images for face recognition. *Proceedings of SPIE*, 8371:1K.1–1K.13, May 2012.
- [51] O. Arandjelovic, R. Hammoud, and R. Cipolla. On person authentication by fusing visual and thermal face biometrics. In *Proceedings of 2006 IEEE*

- International Conference on Video and Signal Based Surveillance*, pages 50–56, November 2006.
- [52] M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu. Fusion of wavelet coefficients from visual and thermal face images for human face recognition — A comparative study. *International Journal of Image Processing*, 4(1):12–23, March 2010.
- [53] M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu. Quotient based multiresolution image fusion of thermal and visual images using Daubechies wavelet transform for human face recognition. *International Journal of Computer Science Issues*, 7(3):18–27, May 2010.
- [54] M. K. Bhowmik, D. Bhattacharjee, D. K. Basu, and M. Nasipuri. Polar fusion technique analysis for evaluating the performances of image fusion of thermal and visual images for human face recognition. In *Proceedings of 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management*, pages 62–69, April 2011.
- [55] 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, February 2007.
- [56] W. K. Wong, J. H. Hui, J. B. M. Desa, N. I. N. B. Ishak, A. B. Sulaiman, and Y. B. M. Nor. Face detection in thermal imaging using head curve geometry. In *Proceedings of the 5th International Congress on Image and Signal Processing*, pages 881–884, October 2012.
- [57] M. Correa, G. Hermosilla, R. Verschae, and J. R. del Solar. Human detection and identification by robots using thermal and visual information in domestic environments. *Journal of Intelligent and Robotic Systems*, 66(1–2):223–243, April 2012.

- [58] J. W. Davis and V. Sharma. Robust background-subtraction for person detection in thermal imagery. In *Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop*, pages 128–135, June 2004.
- [59] F. Suard, A. Rakotomamonjy, and A. Bensrhair. Pedestrian detection using infrared images and histograms of oriented gradients. In *Proceedings of 2006 IEEE Intelligent Vehicles Symposium*, pages 206–212, June 2006.
- [60] L. Zhang, B. Wu, and R. Nevatia. Pedestrian detection in infrared images based on local shape features. In *Proceedings of 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.
- [61] S. Baker and T. Kanade. Hallucinating faces. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 83–88, March 2000.
- [62] C. Liu, H. Y. Shum, and C. S. Zhang. A two step approach to hallucinating faces: Global parametric model and local nonparametric model. In *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 192–198, December 2001.
- [63] H. Huang, H. He, X. Fan, and J. Zhang. Super-resolution of human face image using canonical correlation analysis. *Journal of Pattern Recognition*, 43(7): 2532–2543, July 2010.
- [64] L. An and B. Bhanu. Face image super-resolution using 2D CCA. *Signal Processing*, 103(C):184–194, October 2014.
- [65] X. Ma, J. Zhang, and C. Qi. Hallucinating faces: Global linear modal based super-resolution and position based residue compensation. In *Proceedings of the 15th International Conference on Image Analysis and Processing*, pages 835–843, September 2009.

- [66] W. Liu, D. Lin, and X. Tang. Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation. In *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 478–484, June 2005.
- [67] Z. Hui and K. M. Lam. An efficient local-structure-based face-hallucination method. In *Proceedings of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1265–1268, March 2012.
- [68] P. H. Hennings-Yeomans, S. Baker, and B. V. K. V. Kumar. Simultaneous super-resolution and feature extraction for recognition of low resolution faces. In *Proceedings of 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.
- [69] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. Deblurring face images with exemplars. In *Proceedings of the 13th European Conference on Computer Vision*, pages 47–62, September 2014.
- [70] M. Nishiyama, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi. Facial deblur inference to improve recognition of blurred faces. In *Proceedings of 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1115–1122, June 2009.
- [71] B. Heflin, B. Parks, W. Scheirer, and T. Boult. Single image deblurring for a real-time face recognition system. In *Proceedings of the 36th Annual Conference on IEEE Industrial Electronics Society*, pages 1185–1192, November 2010.
- [72] E. Luo, S. H. Chan, and T. Q. Nguyen. Image denoising by targeted external databases. In *the 39th IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2450–2454, May 2014.
- [73] Y. Tang, R. Salakhutdinov, and G. Hinton. Robust Boltzmann machines for recognition and denoising. In *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2264–2271, June 2012.

- [74] M. Shao, Y. Wang, and Y. Wang. A super-resolution based method to synthesize visual images from near infrared. In *Proceedings of the 16th IEEE International Conference on Image Processing*, pages 2453–2456, November 2009.
- [75] J. Chen, D. Yi, J. Yang, G. Zhao, S. Z. Li, and M. Pietikainen. Learning mappings for face synthesis from near infrared to visual light images. In *Proceedings of 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 156–163, June 2009.
- [76] Z. Zhang, Y. Wang, and Z. Zhang. Face synthesis from near-infrared to visual light via sparse representation. In *Proceedings of 2011 IEEE International Joint Conference on Biometrics*, pages 1–6, October 2011.
- [77] K. Goh, T. Matsukawa, T. Okabe, and Y. Sato. Converting near infrared facial images to visible light images using skin pigment model. In *Proceedings of 2013 IAPR International Conference on Machine Vision Applications*, pages 153–156, May 2013.
- [78] J. Li, P. Hao, C. Zhang, and M. Dou. Hallucinating faces from thermal infrared images. In *Proceedings of 2008 IEEE International Conference on Image Processing*, pages 465–468, October 2008.
- [79] M. Dou, C. Zhang, P. Hao, and J. Li. Converting thermal infrared face images into normal gray-level images. In *Proceedings of the 8th Asian Conference on Computer Vision*, pages 722–732, November 2007.
- [80] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *Advances in Neural Information Processing Systems 3*, pages 572–577, November 1990.
- [81] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *Proceedings of 2003 IEEE Computer Society*

- Conference on Computer Vision and Pattern Recognition Workshops*, pages 53–59, June 2003.
- [82] M. J. Lyons, J. Budynek, A. Plante, and S. Akamatsu. Classifying facial attributes using a 2-D Gabor wavelet representation and discriminant analysis. In *Proceedings of the 4th IEEE Conference on Automatic Face and Gesture Recognition*, pages 202–207, March 2000.
- [83] J. Joo, S. Wang, and S. C. Zhu. Human attribute recognition by rich appearance dictionary. In *Proceedings of the 14th IEEE International Conference on Computer Vision*, pages 721–728, December 2013.
- [84] C. Shan, S. Gong, and P. W. McOwan. Robust facial expression recognition using local binary patterns. In *Proceedings of 2005 IEEE International Conference on Image Processing*, volume 2, pages II-370–II-373, September 2005.
- [85] X. Feng, M. Pietikäinen, and A. Hadid. Facial expression recognition based on local binary patterns. *Pattern Recognition and Image Analysis*, 17(4):592–598, December 2007.
- [86] Z. Ying, L. Cai, J. Gan, and S. He. Facial expression recognition with local binary pattern and Laplacian eigenmaps. In *Emerging Intelligent Computing Technology and Applications: 5th International Conference on Intelligent Computing*, pages 228–235. Springer, September 2009.
- [87] K. Mase. Recognition of facial expression from optical flow. *IEICE Transactions on Information and Systems*, E74-D(10):3474–3483, October 1991.
- [88] V. V. Nabiyev and F. Bölükbaş. Race recognition with local binary pattern. In *Proceedings of 2009 International Conference on Application of Information and Communication Technologies*, pages 1–4, October 2009.
- [89] S. Gutta, H. Wechsler, and P. J. Phillips. Gender and ethnic classification of face images. In *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pages 194–199, April 1998.

- [90] G. Shakhnarovich, P. A. Viola, and B. Moghaddam. A unified learning framework for real time face detection and classification. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 14–21, May 2002.
- [91] B. Moghaddam and M.-H. Yang. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):707–711, May 2002.
- [92] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, September 1995.
- [93] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs. A study of face recognition as people age. In *Proceedings of the 11th IEEE International Conference on Computer Vision*, pages 1–8, October 2007.
- [94] G. Sharma, F. Jurie, and C. Schmid. Expanded parts model for human attribute and action recognition in still images. In *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–659, June 2013.
- [95] G. Christie, A. Parkash, U. Krothapalli, and D. Parikh. Predicting user annoyance using visual attributes. In *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3630–3637, June 2014.
- [96] L. Bourdev, S. Maji, and J. Malik. Describing people: A poselet-based approach to attribute classification. In *Proceedings of 2011 International Conference on Computer Vision*, pages 1543–1550, November 2011.
- [97] Y. Deng, P. Luo, C. C. Loy, and X. Tang. Pedestrian attribute recognition at far distance. In *Proceedings of the 22nd ACM International Conference on Multimedia*, pages 789–792, November 2014.
- [98] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *Proceedings of the 12th IEEE International Conference on Computer Vision*, pages 365–372, September 2009.

- [99] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1962–1977, October 2011.
- [100] Y. Ijiri, S. Lao, T. X. Han, and H. Murase. Efficient facial attribute recognition with a spatial codebook. In *Proceedings of the 20th IAPR International Conference on Pattern Recognition*, pages 1461–1464, August 2010.
- [101] J. Thornton, J. Baran-Gale, D. Butler, M. Chan, and H. Zwahlen. Person attribute search for large-area video surveillance. In *Proceedings of 2011 IEEE International Conference on Technologies for Homeland Security*, pages 55–61, November 2011.
- [102] R. Layne, T. M. Hospedales, and S. Gong. Person re-identification by attributes. In *Proceedings of the 23rd British Machine Vision Conference*, pages 24.1–24.11, September 2012.
- [103] S. Khamis, C.-H. Kuo, V. K. Singh, V. D. Shet, and L. S. Davis. Joint learning for attribute-consistent person re-identification. In *Proceedings of the 13th European Conference on Computer Vision*, pages 134–146, September 2014.
- [104] A. Li, L. Liu, K. Wang, S. Liu, and S. Yan. Clothing attributes assisted person reidentification. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(5):869–878, May 2015.
- [105] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, and L. Bourdev. PANDA: Pose aligned networks for deep attribute modeling. In *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1644, June 2014.
- [106] S. Kang, D. Lee, and C. D. Yoo. Face attribute classification using attribute-aware correlation map and gated convolutional neural networks. In *Proceedings of 2015 IEEE International Conference on Image Processing*, pages 4922–4926, September 2015.

- [107] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez. Automatic features localization in thermal images for facial expression recognition. In *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 14–21, June 2005.
- [108] B. Hernandez, G. Olague, R. Hammoud, L. Trujillo, and E. Romero. Visual learning of texture descriptors for facial expression recognition in thermal imagery. *Computer Vision and Image Understanding*, 106(2–3):258–269, May–June 2007.
- [109] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages I-511–I-518, December 2001.
- [110] Paul Viola and Michael J Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, May 2004.
- [111] W. Härdle and L. Simar. Canonical correlation analysis. In *Applied Multivariate Statistical Analysis*, pages 361–372. Springer, April 2007.
- [112] L. Saul and S. Roweis. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 4:119–155, December 2003.
- [113] Thermal video system advanced thermo TVS-500EX. <http://www.infrared.avio.co.jp/en/products/ir-thermo/lineup/tvs-500ex/spec.html>. Accessed July 7th, 2017.
- [114] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13):800–801, June 2008.
- [115] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

- [116] M. Turk and A. Pentland. Eigenfaces for recognition. *Cognitive Neuroscience*, 3(1):71–86, 1991.
- [117] P. L. Lai and C. Fyfe. Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems*, 10(5):365–377, October 2000.
- [118] S. Akaho. A kernel method for canonical correlation analysis. In *Proceedings of the 12th International Meeting of the Psychometric Society*, pages 614–619, July 2001.
- [119] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, December 2004.
- [120] D. R. Hardoon, J. Mourão-Miranda, M. Brammer, and J. Shawe-Taylor. Unsupervised analysis of fMRI data using kernel canonical correlation. *NeuroImage*, 37(4):1250–1259, October 2007.
- [121] S.-Y. Huang, M.-H. Lee, and C. K. Hsiao. Nonlinear measures of association with kernel canonical correlation analysis and applications. *Journal of Statistical Planning and Inference*, 139(7):2162–2174, July 2009.
- [122] G. Andrew, R. Arora, J. Bilmes, and K. Livescu. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning*, pages 1247–1255, June 2013.
- [123] T. Bouwmans and E.-H. Zahzah. Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, May 2014.
- [124] C. Guyon, E.-H. Zahzah, and T. Bouwmans. Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis. In *Principal Component Analysis*, pages 223–228. InTech, March 2012.

- [125] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in Neural Information Processing Systems 22*, pages 2080–2088, December 2009.
- [126] B. Wohlberg, R. Chartrand, and J. Theiler. Local principal component pursuit for nonlinear datasets. In *Proceedings of the 37th IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3925–3928, March 2012.
- [127] H. Xu, C. Caramanis, and S. Sanghavi. Robust PCA via outlier pursuit. In *Advances in Neural Information Processing Systems 23*, pages 2496–2504, December 2010.
- [128] T. Zhou and D. Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *Proceedings of the 28th International Conference on Machine Learning*, pages 33–40, June 2011.
- [129] X. Ding, L. He, and L. Carin. Bayesian robust principal component analysis. *IEEE Transactions on Image Processing*, 20(12):3419–3430, December 2011.
- [130] C. Guyon, T. Bouwmans, and E.-H. Zahzah. Foreground detection via robust low rank matrix decomposition including spatio-temporal constraint. In *Proceedings of the 11th Asian Conference on Computer Vision Workshops*, pages 315–320, November 2012.
- [131] C. Guyon, T. Bouwmans, and E.-H. Zahzah. Foreground detection via robust low rank matrix factorization including spatial constraint with iterative reweighted regression. In *Proceedings of the 21st IAPR International Conference on Pattern Recognition*, pages 2805–2808, November 2012.
- [132] C. Guyon, T. Bouwmans, and E.-H. Zahzah. Moving object detection via robust low rank matrix decomposition with IRLS scheme. In *Proceedings of the 8th International Symposium on Visual Computing*, pages 665–674, July 2012.

- 
- [133] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos. Sparse Bayesian methods for low-rank matrix estimation. *IEEE Transactions on Signal Processing*, 60(8):3964–3977, August 2012.
- [134] C. A. Wunderlich and E. Seguin. *Medical Thermometry, and Human Temperature*. William Wood & Company, 1871.
- [135] G. Kelly. Body temperature variability (Part 1): A review of the history of body temperature and its variability due to site selection, biological rhythms, fitness, and aging. *Alternative Medicine Review*, 11(4):278–293, December 2006.
- [136] G. Kelly. Body temperature variability (Part 2): Masking influences of body temperature variability and a review of body temperature variability in disease. *Alternative Medicine Review*, 12(1):49–62, March 2007.
- [137] M. Sund-Levander, C. Forsberg, and L. K. Wahren. Normal oral, rectal, tympanic and axillary body temperature in adult men and women: A systematic literature review. *Scandinavian Journal of Caring Sciences*, 16(2):122–128, June 2002.



# List of Publications

## Journal

- [1] B. Kresnaraman, Y. Kawanishi, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase. Human wearable attribute recognition using probability-map-based decomposition of thermal infrared images. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E100-A(3): 854–864, March 2017.
- [2] B. Kresnaraman, Y. Kawanishi, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase. Reconstructing face image from the thermal infrared spectrum to the visible spectrum. *Sensors*, 16(4): 1–6, April 2016.
- [3] B. Kusumoputro, S. R. Imantaka, Lina, and B. Kresnaraman. Face recognition system of infra-red images using ensemble back-propagation neural networks. *Journal of Artificial Intelligence*, 7(A11): 401–426, April 2011.
- [4] B. Kusumoputro, Lina, and B. Kresnaraman. Improvement of recognition capability of fuzzy-neuro LVQ using fuzzy eigen decomposition for discriminating three-mixture fragrances odor. *Information Technology Journal*, 10(12): 2385–2391, 2011.

## International Conference

- [5] B. Kresnaraman, Y. Kawanishi, T. Takahashi, D. Deguchi, Y. Mekada, I. Ide, and H. Murase. Headgear recognition by decomposing human images in the

- thermal infrared spectrum. In *Proceedings of the 15th International Conference on Quality in Research*, July 2017.
- [6] B. Kresnaraman, Y. Kawanishi, T. Takahashi, D. Deguchi, Y. Mekada, I. Ide, and H. Murase. Human wearable attribute recognition using decomposition of thermal infrared images. In *Proceedings of the 22nd Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pages 123–127, February 2016.
- [7] B. Kresnaraman, T. Takahashi, D. Deguchi, Y. Mekada, I. Ide, and H. Murase. Learning based reconstruction of grayscale face image from far-infrared image. In *Proceedings of the 2nd IAPR Asian Conference on Pattern Recognition*, pages 435–439, November 2013.
- [8] B. Kresnaraman, T. Yoshida, K. Doman, T. Takahashi, D. Deguchi, Y. Mekada, and H. Murase. Estimating grayscale face image from thermal image using canonical correlation analysis. In *Proceedings of 2013 International Workshop on Advanced Image Technology*, page 781, January 2013.
- [9] B. Kusumoputro, B. Kresnaraman, and Lina. Development of fuzzy manifold and fuzzy nearest distance for pose estimation of degraded face images. In *Proceedings of the 9th WSEAS International Conference on Applications of Computer Engineering*, pages 180–185, March 2010.

## Domestic Conference

- [10] B. Kresnaraman, Y. Kawanishi, T. Takahashi, D. Deguchi, Y. Mekada, I. Ide, and H. Murase. A preliminary study on human attribute classification in thermal image. In *Proceedings of 2015 IEICE General Conference*, vol. Info. & Sys. 2, page 66, March 2015.

## Award

- [1] Best Student Presentation Award in *the 15th International Conference on Quality in Research*, July 2017.

- [2] Excellent Paper Award in *the 22nd Korea-Japan Joint Workshop on Frontiers of Computer Vision*, February 2016.