INVESTIGATIONS ON MULTI-SENSOR IMAGE SYSTEM AND ITS SURVEILLANCE APPLICATIONS

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the PhD degree in Electrical Engineering

The Ottawa-Carleton Institute for Electrical and Computer Engineering
Faculty of Engineering
University of Ottawa

Zheng Liu

Ottawa, Canada, September 2007

Abstract

THIS thesis addresses the issues of multi-sensor image systems and its surveillance applications. The advanced surveillance systems incorporate multiple imaging modalities for an improved and more reliable performance under various conditions. The so-called image fusion technique plays an important role to process multi-modal images. The use of image fusion techniques has been found in a wide range of applications. The fusion operation is to integrate features from multiple inputs into the fused result.

The image fusion process consists of four basic steps, i.e. preprocessing, registration, fusion, and post-processing or evaluation. This thesis focuses on the last three topics. The first topic is the image registration or alignment, which is to associate corresponding pixels in multiple images to the same physical point in the scene. The registration of infrared and electro-optic video sequences is investigated in this study. The initial registration parameters are derived from the match of head top points across the consecutive video frames. Further refinement is implemented with the maximum mutual information approach. Instead of doing the foreground detection, the frame difference, from which the head top point is detected, is found with an image structural similarity measurement.

The second topic is the implementation of pixel-level fusion. In this study, a modified fusion algorithm is proposed to achieve context enhancement through fusing infrared and visual images or video sequences. Current available solutions include adaptive enhancement and direct pixel-level fusion. However, the adaptive enhancement algorithm should

be tuned to the specific images manually and the performance may not always satisfy the application. Direct fusion of infrared and visual images does combine the features exhibiting in different ranges of electromagnetic spectrum, but such features are not optimal to human perception. Motivated by the adaptive enhancement, a modified fusion scheme is proposed. The visual image is first enhanced with the corresponding infrared image. Then, the enhanced image is fused with the visual image again to highlight the background features. This achieves a context enhancement most suitable for human perception.

As the application of multi-sensor concealed weapon detection (CWD) is concerned, this thesis clarifies the requirements and concepts for CWD. How the CWD application can benefit from multi-sensor fusion is identified and a framework of multi-sensor CWD is proposed. A solution to synthesize a composite image from infrared and visual image is presented with experimental results. The synthesized image, on one hand provides both the information of personal identification and the suspicious region of concealed weapons; on the other hand implements the privacy protection, which appears to be an important aspect of the CWD process.

The third topic is about the fusion performance assessment. So far a number of fusion algorithms have been and are being proposed. However, there is not such a solution to objectively assess those fusion algorithms based on how the features are fused together. In this study, the evaluation metrics are developed for reference-based assessment and blind assessment respectively. An absolute measurement of image features, namely phase congruency, is employed.

This thesis only addresses a limited number of closely related issues regarding to the multi-sensor imaging systems. It is definitely worth further investigations on these topics as discussed in the conclusion of this thesis. In addition, future work should include the reliability and optimization study of multiple image sensors from applications' and human perception-related perspectives. This thesis could be a contribution to such research.

To my grandparents, parents, wife, beloved families, and friends.

Acknowledgments

MY first thank should go to my thesis supervisor Dr. Robert Laganière. When I decided to gain more knowledge in computer vision, Dr. Laganière offered me such a chance to explore what attracted me most. During the time I earned the course credits, passed the comprehensive exam, and did the research work for the thesis, he showed great patience on supervising. The thesis is definitely a result of his hard work. I really enjoyed the study at the University of Ottawa and have benefited from the discussion and collaboration with the other students and professors. Their sincere help and suggestions do contribute to this study and improvement of the thesis.

I want to express my appreciation to Dr. Koichi Hanasaki, who taught me how to observe, research, and analyze when I studied in Japan. The experience gained during that time is the greatest treasure to me. Mr. David S. Forsyth is also appreciated for his open mind and valuable support to the study presented in this thesis.

Finally, I would like to express my great love to my families. Without their supports, patience, and love, I cannot imagine how I can accomplish the work that interests me.

Contents

Appendix A

Appendix B

Appendix C

List of Tables

List of Figures

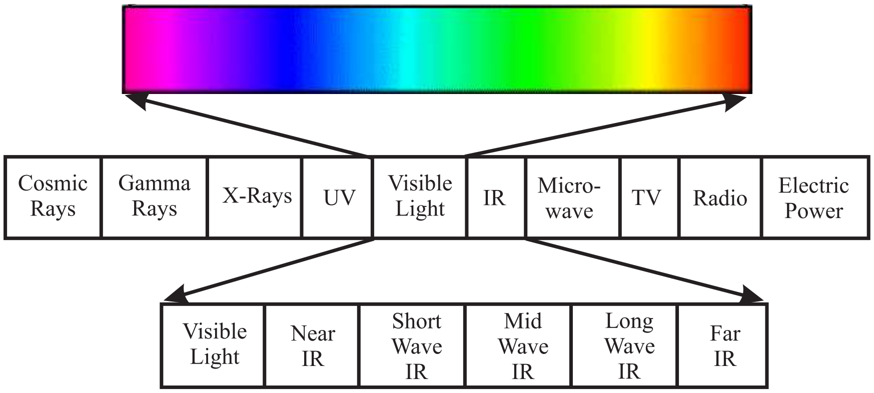
Chapter 1

Introduction

1.1 Motivation and Objective

WITH the development of imaging sensors, it is possible for heterogeneous image modalities to perform across different wavebands of the electromagnetic spectrum [2, 4]. The information acquired from these wavebands can be combined with a so-called image fusion technique, in which an enhanced single view of a scene with extended information content is achieved as the final result. The application of image fusion techniques can be found in a wide range of applications including multi-focus imagery, concealed weapon detection (CWD), intelligent robots, surveillance systems, medical diagnosis, remote sensing, non-destructive testing (NDT), etc.[5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16].

All the possible electromagnetic radiation consists of the electromagnetic spectrum as shown in Figure 1.1(a) and corresponding wavelengths are listed in Table 1.1. The wavelength of the visible light ranges approximately from 390 *nm* to 770 *nm*. After the visible light comes the infrared (IR), which ranges from 770 *nm* to 1 *mm* and is further divided into five parts, e.g. near IR, short IR, mid-wave IR, long-wave IR, and far IR.



(a) The whole electromagnetic spectrum.

Figure 1.1: The electromagnetic spectrum [2].

Table 1.1: The electromagnetic wavelength table [1]

Electromagnetic Wave	Wavelength $\lambda(\mu m)$
Cosmic Rays	$\lambda < 10^{-7}$
Gamma Rays	$10^{-4} > \lambda > 10^{-8}$
X-Rays	$0.1 > \lambda > 10^{-7}$
UV	$0.39 > \lambda > 0.01$
Visible Light	$0.77 > \lambda > 0.39$
IR	$10^{3} > \lambda > 0.77$
Microwave	$10^{6} > \lambda > 10^{3}$
TV and Radio Wave	$10^{11} > \lambda > 10^{6}$
Electric Power	$\lambda > 10^{10}$

Objects having temperature more than $0K$ $(-273.15°)$can generally emit infrared radiation across a spectrum of wavelengths. The intensity of an object's emitted IR energy is proportion to its temperature. The emitted energy measured as the target's emissivity, which is the ratio between the emitted energy and the incident energy, indicates an object's temperature. At any given temperature and wavelength, there is a maximum amount of radiation that any surface can emit. If a surface emits this maximum amount of radiation, it is known as a blackbody. Planck's law for blackbody defines the radiation as [17]:

$$I_{\lambda,b}\left(\lambda,T\right)=\frac{2hc^{2}}{\lambda^{5}}\frac{1}{e^{\frac{hc}{\lambda kT}}-1} \tag{1.1}$$

where $I\left(\lambda,T\right)$ is the spectral radiance or energy per unit time, surface area, solid angle, and wavelength (Uit: $W\ m^{2}\ \mu m^{-1}\ sr^{-1}$). The meaning of each symbol in above equation is listed below [18]:

$$\begin{aligned} \lambda &: wavelength\ (meter) \\ T &: Temperature\ (kelvin) \\ h &: Planck's\ constant\ (joule/hertz) \\ c &: speed\ of\ light\ (meter/second) \\ k &: Boltzmann's\ constant\ (joule/kelvin) \end{aligned}$$

Usually, objects are not blackbodies. According to Kirchhoff's law, there is $R+\epsilon=1$, where ϵ is the emissivity and R is the reflectivity. Emissivity is used to quantify the energy-emitting characteristics of different materials and surfaces. The emitted energy of an object reaches the IR sensor and is converted into an electrical signal. This signal can be further converted into a temperature value based on the sensor's calibration equation and the object's emissivity. The signal can be displayed and presented to the end users. Thus, thermography can "see in the night" without an infrared illumination. The amount of radiation increases with temperature; therefore, the variations in temperature can be identified by thermal imaging. The IR cameras can generally be categorized into two types: cooled infrared detectors and uncooled infrared detectors. They can detect the difference in infrared radiation with insufficient illumination or even in total darkness. The use of thermal vision techniques can be found in numerous applications such as military, law enforcement, surveillance, navigation, security, and wildlife observation [19]. The IR image can provide an enhanced spectral range that is imperceptible to human beings and contribute to the contrast between objects of high temperature variance and the environment. Compared with a visual image, the IR image is represented with a different intensity map. The same scene exhibits different features existing in different electromagnetic spectrum bands.

The purpose of this study is to investigate how the information captured by multiple imaging systems can be combined to achieve an improved understanding or awareness of the situation. This thesis will focus on the registration and fusion of IR and visual images in surveillance applications and the the fusion performance assessment issue.

1.1.1 The Statement of Problems

The procedure of fusing multi-modal images is depicted in Figure 1.2. There are basically four major steps, i.e. pre-processing, registration, fusion, and post-processing. In the preprocessing stage, a filtering operation can be applied to remove the noises introduced during the image acquisition process. The registration is to align corresponding pixels associated with the same physical points in the real world[1]. Then, the registered images are combined with the fusion algorithms, which can be implemented at three different levels, i.e. pixel

[1] We assume that the images have been temporally synchronized.

level, feature level, and symbol level. The fused result can be presented to the end user or for further analysis, depending on the requirements of the application. The question is “what is the most appropriate solution to a specific application?”.

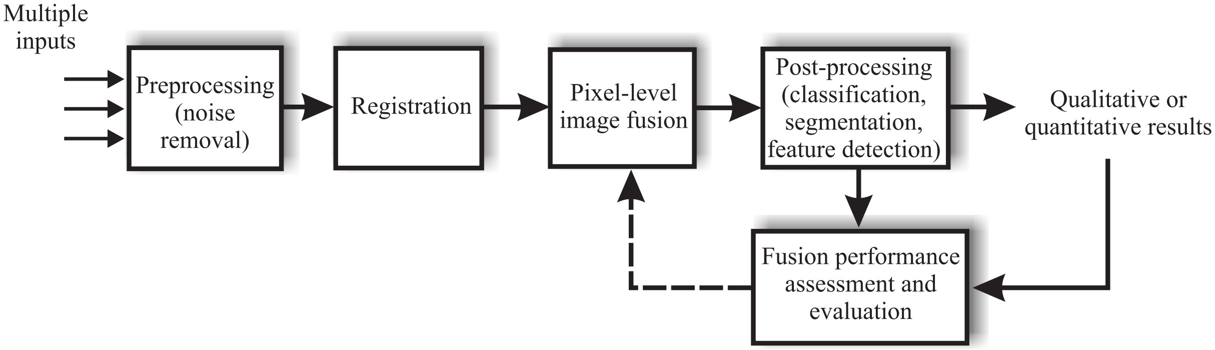

Multiple inputs
Preprocessing (noise removal)
Registration
Pixel-level image fusion
Post-processing (classification, segmentation, feature detection)
Qualitative or quantitative results
Fusion performance assessment and evaluation

Figure 1.2: The procedure for multi-modal image fusion.

Obtaining a fused result does not come to the end of the fusion process. Another challenge is the assessment of the fused result. Again, this is a typically application-dependent issue. The question could be “what is expected from the fusion output?” and “what is the metric to assess the fusion result?”. If there is a perfect reference, the fused image can be compared with this reference directly. However, this is not the case in most applications, i.e. no such perfect reference is available all the time. We still need to come up with an evaluation metric, either subjective or objective, to evaluate the fusion result. Moreover, if the assessment metric is properly used to guide the fusion process, adaptive fusion can be implemented.

There are problems associated with each step during the whole fusion process and those issues have not been fully explored and addressed so far. This thesis will focus on three major problems: registration, fusion, and objective assessment. Registration refers to the alignment of multi-modal images so that pixel-based operations can be applied. Image fusion should meet the target of the specific applications. The quantitative metric will tell how well those fusion algorithms perform.

1.1.2 Objectives of the Research

The objectives of the research are twofold. The first objective of this study is to develop fusion algorithms for the infrared and visual images in surveillance applications. The studies include the registration of IR and visual video sequences, multi-modal image (video sequence) enhancement, and concealed weapon detection and visualization. Before any fusion operation, the pixels from multi-modal images should be aligned through the registration process so that pixel-level processing can be applied directly. The study on the enhancement from multi-modal images will investigate how the information can be fused to achieve a better fidelity. When the IR camera is employed for the application of concealed weapon detection, the study is to investigate how the image fusion technique can facilitate the security screening process and system performance enhancement.

The second is to develop an objective assessment scheme for generic image fusion applications, more precisely, a feature-based quantitative evaluation of combinative pixel-level image fusion. Herein, the purpose of fusion is to combine the complementary features from multiple input images and the fused result is still an image. The fusion result is often subjectively assessed. However, an objective assessment is desired to get a full understanding and more accurate evaluation of the fusion process. The second objective is to develop such a quantitative metric, which is based on the measure of image features. The image fusion will benefit from this metric in two aspects. On one hand, we can select the most appropriate fusion algorithm for a specific application (image modality) based on the metric;

on the other hand, the metric can be employed as guidance to supervise the fusion process [20] and thus an adaptive fusion process can be implemented to achieve an optimized fusion result.

As illustrated with Figure 1.2, these two aspects consist of the essentials of a fusion process and are closely related to each other. In order to highlight the contribution of the the studies, the two aspects are treated as two major objectives in this thesis.

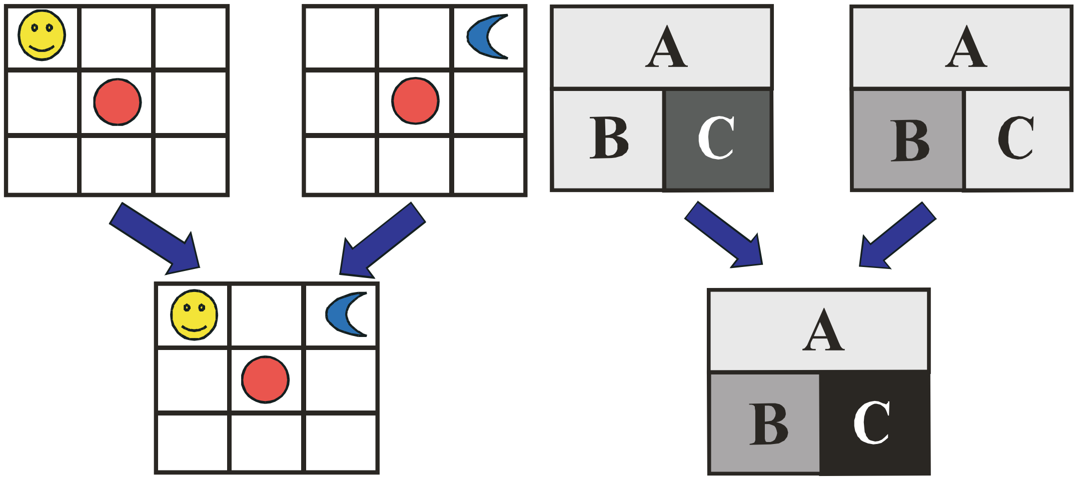
1.2 Background and Significance

1.2.1 Multi-Sensor Image System

Processing of Multi-Sensor Images

One of the most important processing for multi-sensor images is the so-called image fusion technique. This terminology derives from a more general concept “sensor fusion”, which is also known as “data fusion” or “information fusion”. The definition of data fusion may vary with people and their perspectives. The one proposed by Wald is “*Data fusion is a formal framework in which are expressed the means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application.*” [21]. The implementation of data fusion can be achieved at three levels, e.g. signal level, feature (attribute) level, and symbol (decision) level. Pixel-level fusion refers to the fusion of multiple images at the signal level, i.e. an image-in image-out process. As illustrated in Figure 1.2, when pixel-level fusion is considered, a preprocessing operation needs to be carried out to remove the noises introduced by image sensors or cameras.

The purpose of image fusion is to generate a composite image from multiple inputs.



(a) Feature integration.

(b) Feature discrimination.

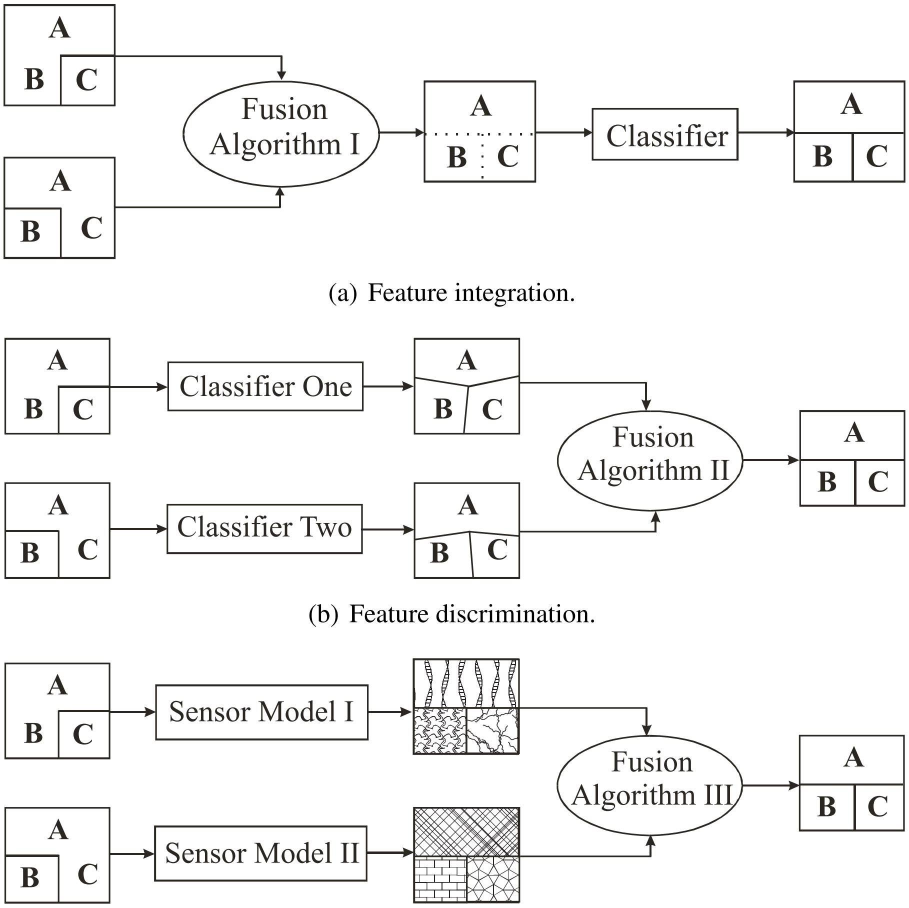
Figure 1.3: Two aspects of the image fusion problem.

The fused image can provide more complete information about the scene than any individual images. The fusion of multi-source images can also be implemented at the previously mentioned three levels and the fusion process depends on the specific application. The image fusion problems can be classified into two categories as shown in Figure 1.3: one is for feature integration; the other is for feature classification or characterization. In Figure 1.3(a), different features (face and moon) from two images can be integrated in a new image. The circle in the center can be some common features from the input images. More information about the scene is achieved through the fusion at pixel level. Both the face and moon are available in the fused image. However, this does not reach the end of analysis. Those features can be further identified at the post-processing stage. In the second case shown in Figure 1.3(b), sensor one discriminates A and B from C while sensor two discriminates A and C from B. Herein, the letter A, B, and C does not necessarily indicate the objects only; they may have a more extensive meaning like image feature. The goal of fusing image one and two is to distinguish A, B, and C from each other. Usually, such fusion is implemented at a decision level and the fused image is a thematic map of the scene. This type of fusion often finds applications in the fields of remote sensing and medical diagnosis.

Image Fusion: the Concept and Implementation

As stated in the previous section, image fusion is one of the most important techniques to process multi-sensor images. Three types of fusion schemes are given in Figure 1.4 [22, 15], where the letter A, B, and C represent different image features. The first type of image fusion is to combine the salient features from the two images as indicated in Figure 1.4(a). The image enhancement and denoising will be achieved at the pixel-level fusion. The feature A, B, and C should be much easier to identify from the fused image. Further processing, e.g. classifying or quantifying process, is needed to fully discriminate the fusion output. The approaches to accomplish this type of fusion process include multiresolution analysis (MRA), optimization, and heuristic methods. The second and third types of image fusion in Figure 1.4(b) and 1.4(c) are implemented at a higher level, i.e. feature or decision level. In the second scheme, the pre-processing (classification) units are designed for the input images. The units can be a segmenting algorithm, data clustering, neural network model or other types of classifiers. The preliminary results are then fused by probabilistic theory, fuzzy logic operator, or numerical combination to generate a thematic map. This actually implements a classifier fusion for each pixel. Instead of using the categorizing methods, an alternative approach is implemented by modeling the image data. Although a physical model is preferred, it is difficult to achieve. Thus, statistical methods can be used to build the image sensor model through calibration, supervised, or unsupervised learning.

The outputs are fused to generate the posterior probabilities by applying Dempster-Shafer theory, Bayesian inference or fuzzy logic methods. The final decision is made based on this value.



(a) Feature integration.

(b) Feature discrimination.

(c) Feature discrimination.

Figure 1.4: The image fusion schemes.

For the feature discrimination, the fusion result is relatively easy to assess by using the classification rate, receiver operating characteristic curve, or/and confusion matrix. For the fusion of feature integration as shown in Figure 1.4(a), the performance of the post-processing could be a good indicator for this. In this thesis, a more general measurement of image feature is considered for the fusion performance assessment.

The study presented in this thesis is closely related to the multiresolution image fusion (MRIF), which implements an image-in-image-out fusion process. The MRIF uses the multiresolution algorithm to represent the image in the transform domain, where image features can be easily accessed and manipulated. The pixel-level fusion is implemented by combining the coefficients, image components, or sub-images. The inverse transform gives the fused results. A brief review of the MRIF can be found in Chapter 3.

1.2.2 Implementation of Image Fusion for Surveillance Applications

Registration of Multi-Modal Video Sequences

Image registration is to assure the corresponding pixels in multiple images associate the same physical points of the scene. The image registration itself is worth a monograph to cover the relevant research topics. Zitova and Flusser presented a comprehensive survey of the image registration techniques in [23]. A general registration procedure consists of four steps: feature detection, feature matching, mapping function design, and image transformation and resampling [23]. The IR and visual images exhibit different intensity values for the same scene. Thus, it is difficult to register the images based on the gray-scale characteristics. Various solutions have been proposed for detecting and matching features from the two types of images [24, 25, 26, 27, 28, 29]. This aspect will be discussed in more details in Chapter 2. A literature review is provided in that chapter.

Night Vision for Context Enhancement

Some automotive manufacture, like BMW and GM, offer the night vision system as a high-tech option for their flagship vehicles. After dark, the chances of being in fatal car crash go up sharply, though traffic is way down [30]. The night vision system can help the drivers see as much as three or four times further ahead and quickly distinguish objects [30]. Figure 1.5 gives a snapshot of the night vision system on a BMW vehicle. Similarly, for a surveillance application the objective may be to detect and track human body in the environment with inadequate illumination.

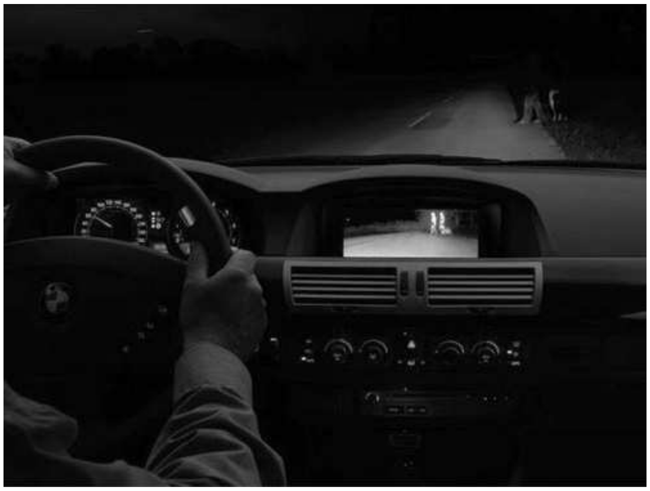


Figure 1.5: The BMW night vision system on a vehicle (courtesy of BMW).

The enhancement of a visual image includes the basic operations on image histogram, like histogram equalization and matching, and operations for adaptive enhancement [31, 32]. Now, the problem becomes "how the processing could be performed when the corresponding IR image is available". The pixel-level image fusion is a well-adopted method [33].

As addressed in previous section, the source images need to be fully registered before a fusion operation is applied. Any enhancement may be applied to the visual image before the fusion operation as proposed by Tao et al. [34]. The purpose of the fusion operation is to highlight the objects with high temperature variance and preserve the details of the background. Chapter 3 proposes solutions to this problem.

Concealed Weapon Detection (CWD) and Visualization

To address the emerging threats from terrorists, there is a need to develop an efficient technique for heightened security requirements and law enforcement. Currently, airport staffs examine passengers with metal detector, hand wands, and physical searches [35] and passengers with concealed objects may not be detected. Imaging systems with a radiation wavelength longer than 20 microns can penetrate clothing and thus have the potential capability to detect concealed weapons [36]. The enabling sensing mechanisms being studied include infrared, acoustic, millimeter wave (MMW), X-ray sensors and so on [6]. Multiple image modalities of different radiation wavelengths can provide complementary information about the scene. However, some imaging techniques, such as X-ray, can provide detailed images of anatomical features. Displaying such features is a violation of individual privacy [37].

It is claimed that the fusion of a MMW image and its corresponding infrared or electro-optical image can achieve more complete information [38]. The infrared (IR) imagers cannot penetrate heavy clothing but operate at a reasonably longer range whereas MMW sensors have a good penetration at a short range [9]. A visual image does not provide any information about the concealed weapons. However, the facial pattern of suspicious people may be available from a visual image. Thus, the fusion of visual image with other image modalities such as MMW image can provide information of both the personal identification and concealed weapons. As a result, the concealed weapon can be easily located in the fused image that is most suitable for human perception. In other words, the composite image can give detailed information about both the person's appearance and hidden weapon. The operator can promptly relate the weapon to the person in crowds who is carrying it without looking at two separate images.

While the detection techniques are approaching an advanced stage, the privacy protection issue comes into view. Fortunately, the fusion of visual image and long-wavelength image will take into account this problem. Therefore, the CWD fusion techniques fall into two categories: one is the fusion for visualization (integration); the other is the fusion for detection (discrimination). There is a simple rule to identify the difference. When the fusion is carried out with a visual image input, this is for visualization. Otherwise, the fusion is for detection. However, these two concepts are not mutually exclusive and the CWD system can be a hybrid one. The visualization is to show the detected weapon. If there is no consideration for the detection, the visualization might not be helpful as expected. Chapter 4 clarifies the image fusion issues related to CWD and proposes an algorithm for the implementation of a CWD system.

1.2.3 Objective Assessment of the Fusion Performance

A diverse range of implementations for image fusion have been proposed. However, the objective assessment of the fusion result still remains a challenge. Sometimes, a subjective evaluation from an expert still plays an important role. When a "perfect" result is available, a straightforward approach is to compare the fused image with the reference image. The commonly used methods include the root mean square error (RMSE), normalized least square error (NLSE), the peak signal to noise ratio (PSNR), correlation (CORR), difference entropy (DE), and mutual information (MI) [39]. In Chapter 5, a feature-based image similarity metric is proposed and applied to assess the fused image through comparing with a reference image. The proposed metric is based on an absolute image feature measurement, namely phase congruency.

The difference between the reference image and the fused one can be calculated and serves as a measurement of the quality of the fused image. The difficulty of the comparison-based approach is that the reference may not be perfect or a reference is not always available in a practical application. Moreover, there is a possibility that images with a similar RMSE value may exhibit a quite different quality [40]. Some other methods that consider human visual system and attempt to incorporate perceptual quality measurement do not always show clear advantage over simple measurements like RMSE and PSNR under the image distortion environments [40, **?**].

Thus, it would be better if the assessment can be accomplished without any reference, i.e. blind assessment, where the fused image only needs to refer to the input images to evaluate itself. The mutual information, image feature, and structural similarity measurement have been used to implement a blind assessment by other researchers [41, 42, 43, 44, 45]. The details of those approaches will be discussed later in Chapter 6.

1.3 Organization of the Thesis

The rest of the thesis is organized as illustrated in Figure 1.6, which clearly shows the relation between these studies. This thesis does not deal with the issues of pre- and post-processing as shown with the blocks in Figure 1.6. Chapter 2 deals with the registration problem for infrared and electro-optic video sequences. This is the first step of the image fusion process. The registration process is a guarantee of the accuracy of pixel-wise operations like pixel-level fusion. In this study, we assume a moving person is present in the captured video sequences. The proposed approach achieves a robust registration using the trajectory of head top points in consecutive frames, which does not rely on the success of other processes such as foreground detection. The head top points are found from the frame difference detected by the image structural similarity measurement. This work provides a basis for further investigation.

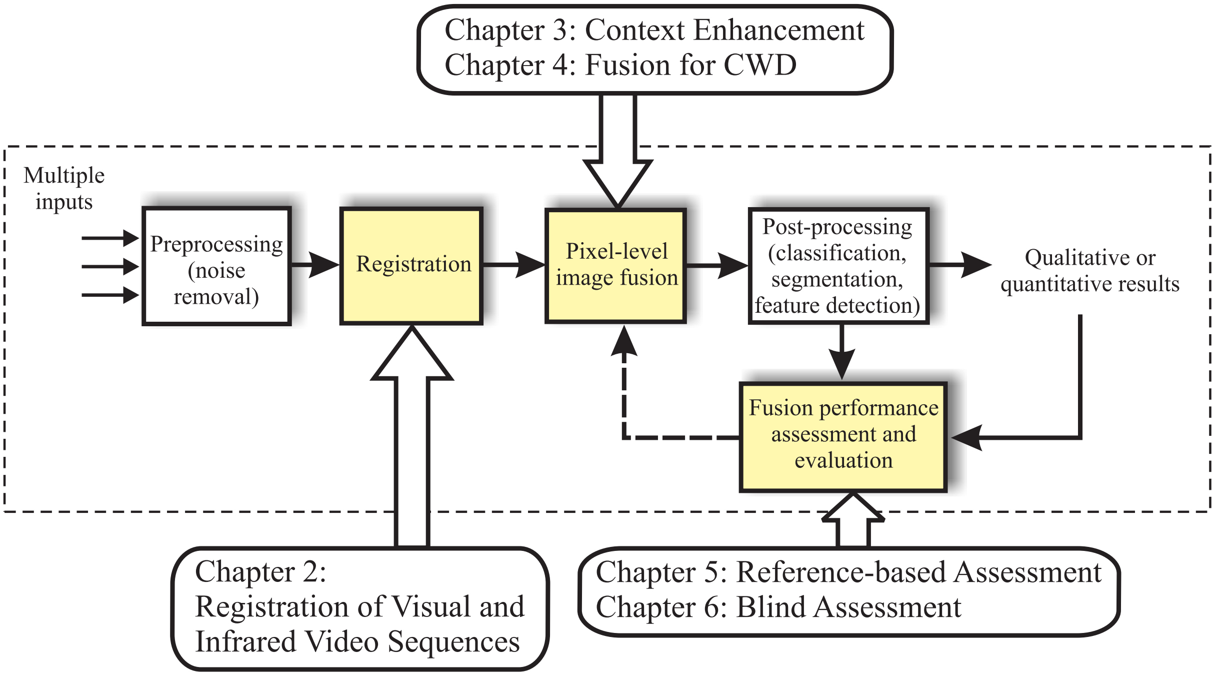

Chapter 3: Context Enhancement
Chapter 4: Fusion for CWD
Multiple inputs
Preprocessing (noise removal)
Registration
Pixel-level image fusion
Post-processing (classification, segmentation, feature detection)
Qualitative or quantitative results
Fusion performance assessment and evaluation
Chapter 2: Registration of Visual and Infrared Video Sequences
Chapter 5: Reference-based Assessment
Chapter 6: Blind Assessment

Figure 1.6: The organization of the thesis.

Context enhancement can be achieved by direct fusion of IR and EO images at pixel level. However, the fused result is not optimal for human perception. Chapter 3 proposes a modified scheme that achieves a better result than the direct fusion and some enhancement approaches. This will help to improve the awareness of the environment and situation under inadequate illumination. The enhanced result is presented in the visual band, which is most suitable for human perception. The method can be applied to surveillance applications as well as vehicular night-vision systems.

Chapter 4 addresses the problems on how the concealed weapon detection application can benefit from multiple imaging modalities. Herein, the direct pixel-level fusion may not be an optimal solution. The experiment is about detecting the region of concealed weapon from IR image and synthesizing the detected region with the EO image. This approach will help protect passengers' privacy and highlight the suspicious region(s) while keeping the information of personal appearance.

To understand the performance of the fusion algorithms, Chapter 5 proposes a reference-based method to evaluate the performance of combinative pixel-level image fusion. The method uses a so-called "phase congruency" measurement as a basis for developing the metric of objective assessment of image similarity. Therefore, the fusion results can be compared with the available reference(s). The proposed metric offers a numeric value indicating the quality of the fused image.

Chapter 6 extends the method in Chapter 5 for the assessment without a reference, which happens in most practical applications. Three implementations are proposed. The first one is based on the modified structural similarity measurement. The second is based on the local correlation of the phase congruency map. The last one considers the principle moments of the phase congruency.

The last chapter (Chapter 7) summarizes the whole thesis and discuss the future potentials and directions for multi-sensor image systems. The Appendices provide the detailed information about the implementation of phase congruency algorithm and experimental results for the comparison of blurred images.

1.4 Contributions of the Thesis

The IR camera has been widely employed in advanced surveillance system. How the IR camera can benefit the whole system through the fusion with corresponding visual frames is not fully understood. One critical step to process multi-modal video sequences is the registration, which assures the accuracy of pixel-wise operations. This study achieved a robust registration, which does not rely on the success of other processes such as foreground detection. The detection of silhouette of difference benefited from the image structural similarity measurement, where a predefined threshold value was applied. Instead of matching multiple features from one IR frame and corresponding visual frame, the proposed method matched the trajectories traced respectively from the silhouettes detected from consecutive frames of IR and visual video sequences. A refinement of the initial estimation was carried out with a maximum mutual information method. This study contributed to the basis for the pixel-level image fusion in the next step.

The studies presented in this thesis facilitate the use of multiple image modalities for the surveillance applications. More precisely, the research investigated how to make use of the information from the infrared and visual images through pixel-level image fusion. However, there is no one-size-fit-all solution. Each application specifies the particular requirements. The fusion algorithms should be tailored or tuned to such a specific scenario, like context enhancement and concealed weapon detection.

The fusion scheme for the image context enhancement improved the awareness of the environment under an inadequate illumination. The fused result was presented in the visual band, which was most suitable for human perception. The infrared and visual images use a different intensity table. The composite image that was fused with the generic multiresolution image fusion method did not resemble the way people perceive the scene with their eyes. The thesis proposed a fusion method to combine the information rather than just pixels from the IR image.

The common understand of image fusion for the concealed weapon detection is somewhat misleading. The research on image fusion for CWD presented in this thesis clarified the role of image fusion. Strictly speaking, the work on image fusion for the concealed weapon “detection” has not been reported so far. Currently available publications by other researchers does not demonstrate the advantages of multi-sensor approaches for the detection of weapons. The capability of certain technique (sensor) is often quantified by the “probability of detection” (POD) results, which indicate the how the sensor performs under varied circumstances, for example, the distance to the object (human body), the thickness of the coat, the environmental temperature, and etc. The requirements for successful detection must be met. The thesis proposes two steps for a general CWD process. The first is the detection process, where the concealed weapon is detected from the IR image. In the second step, the detected “region of interest” (ROI) is embedded in the visual image by the multiresolution image mosaic technique. This clearly identifies the “detection” and “visualization” process. This thesis implemented the fusion of IR and visual image for concealed weapon detection. This work was reported by an American magazine “Bulletin of Atomic Scientist” (issue: April 2006).

Although a lot of image fusion algorithms have been proposed so far, the objective performance evaluation has not been fully explored and addressed. The assessment is an application-dependent process and the evaluation metric may vary with the requirements of the specific application. When the purpose of the fusion is to integrate the features from multi-modal images, such as the context enhancement, there should be a metric that can estimate how the features are fused quantitatively. The purpose of this study is to provide such a solution. This research helps to understand the mechanism of the fusion process and provide the chance to optimize the fusion algorithm based on the evaluation metric. The study developed the methods for reference-based assessment and blind assessment. The proposed evaluation metric will help to understand the effectiveness of the image fusion algorithms for a specific application.

The following is the list of journal publication (either published or submitted) related to this thesis:

- Z. Liu and R. Laganière, “Context Enhancement through Infrared Vision”, *Signal, Image, and Video Processing*, in press (DOI: 10.1007/s11760-007-0025-4).

- Z. Liu, D. S. Forsyth, and R. Laganière, “A Feature-based Metric for the Quantitative Evaluation of Image Fusion”, *Computer Vision and Image Understanding*, in press (DOI:10.1016/j.cviu.2007.04.003).

- J. Y. Zhao, R. Laganière, and Z. Liu, “Performance Assessment of Combinative Pixel-Level Image Fusion based on An Absolute Feature Measurement”, *International Journal of Innovative Computing, Information and Control*, Vol.3, No.6, December 2007.

- Z. Liu and R. Laganière, “The Use of Phase Congruence Measurement for Image Similarity Assessment”, *Pattern Recognition Letters*, Vol.28, pp166-172, 2007.

- Z. Liu, Z. Xue, R. S. Blum, and R. Laganière, “Concealed Weapon Detection and Visualization in a Synthesized Image”, *Pattern Analysis and Applications*, Vol.8, No.4, pp375-389, February 2006.

- Z. Liu, Z. Xue, T. Macuda, D. S. Forsyth, and R. Laganière, “Concealed Weapon Detection: Data Fusion Perspective”, *Journal of Aerospace Computing, Information, and Communication, in review*.

The following paper were refereed and presented on conferences:

- Z. Liu and R. Laganière, “Registration of IR and EO Video Sequence based on Frame Difference”, The First International Workshop on Video Processing and Recognition, Montreal, QC, Canada, 2007.

- J. Y. Zhao, R. Laganière, and Z. Liu, "Image Fusion Algorithm Assessment based on Feature Measurement", International Conference on Innovative Computing, Information and Control, August 30 - September 1, 2006, Beijing, China.

- Z. Liu and R. Laganière, "Image Fusion Algorithm Assessment based on Feature Measurement", IEEE International Conference on Acoustics, Speech, and Signal Processing, May 14-19, 2006, Toulouse, France.

Chapter 2

Registration of Visual and Infrared Video Sequences

2.1 Introduction

THE vision technique beyond visible spectrum becomes essential to advanced surveillance systems. The performance of the systems can be enhanced through taking full advantage of the information captured across the electromagnetic spectrum. This makes the surveillance systems more robust and reliable under different conditions, such as a noisy and cluttered background, poor lighting, smoke, and fog. The technique to achieve this is known as information or sensor fusion. Depending on the requirements, the fusion of multimodal images can be implemented at different levels with varied fusion algorithms [33, 12].

The infrared (IR) camera uses thermal or photonic/quantum detectors to tell the difference in infrared radiation of different objects. The electro-optical (EO) sensors, e.g. CCD or CMOS cameras, capture the reflective light properties of objects [46]. Therefore, the visual and IR imagery may provide the complementary information about the scene [46].

Multiple cues provided by the two imaging modalities can be used to accomplish the tasks of detecting, tracking, and analyzing for the surveillance application. A review of software and hardware considerations for using visible and infrared imagery for surveillance applications was presented in [47] by El-Maadi et al. This paper surveyed the research work carried out under the collaboration of three institutions in Québec City, Canada.

Preceding to any other processing, the EO and IR images from the video sequences should be registered so that the corresponding pixels from the two images are associated with the same physical points in the scene. This ensures the correctness of the pixel- and high-level processing. Nevertheless, some high-level processing does not rely on the accuracy of geometric registration [48, 12] and this is the advantage of the high-level processing.

The registration of the images from IR and electro-optic (CCD) camera can be implemented by fully calibrating the two cameras. Yasuda et al. proposed a calibration procedure in [49], where a grid of electrically heated wires were used. This wire grid appeared as a red wire grid in a color image and a bright wire grid in a thermal image. The calibration is to assure the grid lines in the two camera frames will match. The grid lines detected from the IR image will match the lines extracted from the visual image.

Generally speaking, the image registration procedure consists of four basic steps: feature detection, feature matching, mapping function design, and image transformation and resampling [23]. Li et al. registered multi-sensor images with detected contours [25]. In another publication of Li et al. [24], they used a wavelet-based approach to detect image contours and located feature points on the contours by using local statistics of image intensity. The feature points were matched with a normalized correlation method. A consistency-checking step was applied to eliminate the mismatches. Coirs et al. matched the triangles formed by grouped straight line segments extracted from the IR and EO images [26]. Keller et al. proposed a registration method based on an implicit similarity measure, which was invariant to intensity dissimilarities [29]. The authors claimed that this approach was efficient and suitable to register two images of different qualities. However, the physical correspondences may not be fully detected with matchable contours or lines.

Han et al. suggested using the silhouette of a moving human body to register IR and EO images. They found the silhouette by classifying a pixel as belonging to either foreground or background based on the background Gaussian distribution [27]. The centroid and head top points in two pairs of images were used as control points. A genetic algorithm was employed to minimize the registration error function. In [28], Ye et al. proposed using zero-order statistics to detect moving object in a video sequence. Through tracking the feature points, an iterative registration algorithm was implemented. Xu et al. proposed to use a support vector machine (SVM) to detect pedestrians from IR images in [50]. Related work was also reported by Maes et al. and Chen et al. respectively in [51, 52], where the registration was carried out based on maximizing mutual information of two image regions. However, the images must be roughly registered with some prior knowledge in the surveillance application and the solution was not available in that publication [52].

The moving object detection, which is also known as the background maintenance, still remains a challenge for surveillance applications. For the millimeter wave (MMW) video sequence, such detection can be more difficult due to it blurry nature [52]. In this chapter, a registration method, which uses the silhouette of the frame difference instead of the silhouette of moving objects, is proposed; therefore, this method does not rely on the success of foreground detection and can be applied to any imaging modality. The frame difference can be steadily detected with the image structural similarity measurement. Instead of extracting feature points from one image, the trajectory formed by the head top points in consecutive frames is used for the initial registration. A refining process is implemented based on the maximum mutual information method [52].

The rest of the chapter is organized as follows. The detailed procedure for registration is described in section (2.2). The whole process consists of two steps, i.e. initial registration and parameter refinement. Experimental results are presented in section (2.3). Discussion and conclusion can be found in section (2.4) and (2.5) respectively.

2.2 Registration based on Frame Difference

The proposed registration process can be implemented in two steps. In the first step, the head top points are detected from the silhouette of frame differences. The initial parameters can be estimated by matching the trajectories in IR and EO sequences. The second step is to refine the registration parameter by directly registering two regions of interest with the mutual information maximization method. The only assumption for our method is that only one person is present in the video sequences as a moving object during the registration process, although it is possible to extend the algorithm to deal with multiple points/features from multiple people in one frame.

2.2.1 Image Similarity Measurement

The simplest way to find the difference between two images is the subtraction operation. However, the threshold value may vary with different video clips and needs to be adjusted manually. In this work, we use the structural similarity measurement (SSIM) to detect the difference between consecutive frames.

Wang and Bovik proposed an universal image quality index (UIQI) in [40]. The universal image quality index is based on the evidence that human visual system is highly adapted to structural information and a measurement of the loss of structural information can provide a good approximation of the perceived image distortion. The definition of the UIQI is [40]:

$$Q = \frac{4\sigma_{ab}\mu_a\mu_b}{\left(\sigma_a^2 + \sigma_b^2\right)\left({\mu_a}^2 + {\mu_b}^2\right)} = \frac{\sigma_{ab}}{\sigma_a\sigma_b} \cdot \frac{2\mu_a\mu_b}{{\mu_a}^2 + {\mu_b}^2} \cdot \frac{2\sigma_a\sigma_b}{\sigma_a^2 + \sigma_b^2} \tag{2.1}$$

where μ_a and μ_b are the average values of image $a(x, y)$ and $b(x, y)$, σ_a , σ_b, and σ_{ab} are the variance and covariances respectively. There are:

$$\mu_a = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} a\left(x, y\right) \tag{2.2}$$

$$\mu_b = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} b\left(x, y\right) \tag{2.3}$$

$$\sigma_a^2 = \frac{1}{MN-1} \sum_{x=1}^{M} \sum_{y=1}^{N} \left(a\left(x, y\right) - \mu_a\right)^2 \tag{2.4}$$

$$\sigma_b^2 = \frac{1}{MN-1} \sum_{x=1}^{M} \sum_{y=1}^{N} \left(b\left(x, y\right) - \mu_b\right)^2 \tag{2.5}$$

$$\sigma_{ab} = \frac{1}{MN-1} \sum_{x=1}^{M} \sum_{y=1}^{N} \left(a\left(x, y\right) - \mu_a\right)\left(b\left(x, y\right) - \mu_b\right) \tag{2.6}$$

This equation has been modified to produce the structural similarity index measure (SSIM), which is better adapted to more general conditions [53]. The SSIM is defined as [53]:

$$SSIM\left(a, b\right) = \left[l\left(a, b\right)\right]^{\alpha}\left[c\left(a, b\right)\right]^{\beta}\left[s\left(a, b\right)\right]^{\gamma} \tag{2.7}$$

where there are:

$$l\left(a,b\right) = \frac{2\mu_a\mu_b + C_1}{\mu_a^2 + \mu_b^2 + C_1} \tag{2.8}$$

$$c\left(a,b\right) = \frac{2\sigma_a\sigma_b + C_2}{\sigma_a^2 + \sigma_b^2 + C_2} \tag{2.9}$$

$$s\left(a,b\right) = \frac{\sigma_{ab} + C_3}{\sigma_a\sigma_b + C_3} \tag{2.10}$$

In equation (2.7), three components are clearly defined to measure the degree of linear correlation between image a and b. The first one, $l(a,b)$, measures how the mean luminance is between the two images while the second $c(a,b)$ estimates the contrast. The third one $s(a,b)$ is the correlation of structure. The parameter α, β, and γ can be used to adjust the relative importance of the three components.

By setting $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, equation (2.7) becomes:

$$SSIM\left(a,b\right) = \frac{\left(2\mu_a\mu_b + C_1\right)\left(2\sigma_{ab} + C_2\right)}{\left(\mu_a^2 + \mu_b^2 + C_1\right)\left(\sigma_a^2 + \sigma_b^2 + C_2\right)} \tag{2.11}$$

In above equation, two constant values C_1 and C_2 are defined to avoid the instability when the denominators are very close to zero. These two values are further determined by two subjectively selected values K_1, K_2, and the dynamic range of the pixel values, i.e. $C_1 = \left(K_1L\right)^2$ and $C_2 = \left(K_2L\right)^2$. For a 8-bit gray-scale image, L is selected as 255.

An example of applying SSIM to find the frame difference is given in Figure 2.1. The SSIM maps are generated from two adjacent frames for IR and EO sequences respectively. The mean value of the SSIM map gives an index value, which indicates how different the two images are. In our application, we use the SSIM map instead.

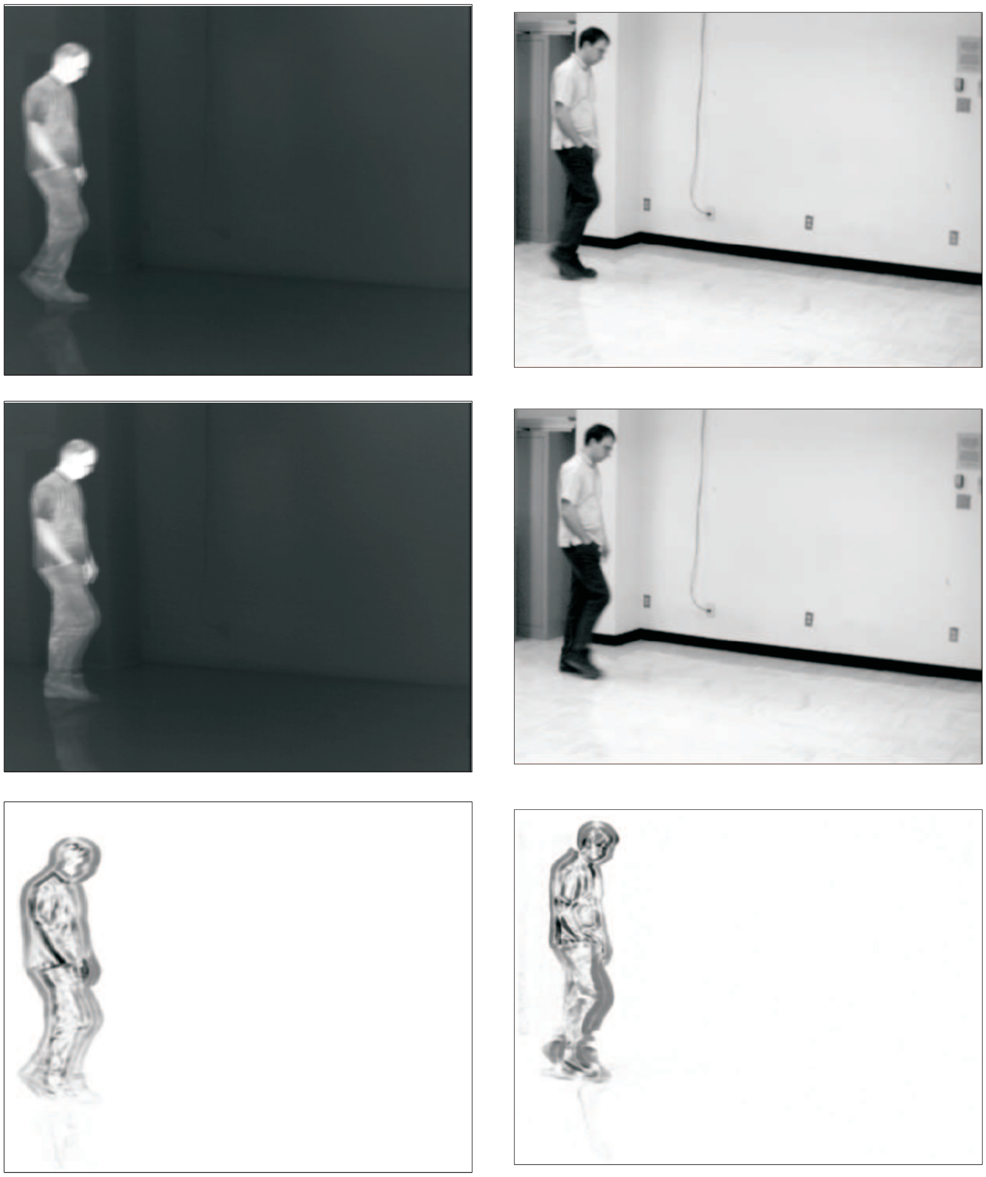


Figure 2.1: The example of SSIM. On the left column are the IR images. Right column is from EO camera. Two adjacent frames and their SSIM map are from the top to bottom.

2.2.2 Silhouette Extraction

Once the SSIM maps are obtained. The detection of the frame difference is straightforward. Simply applying a fixed threshold value to both SSIM maps, two binary images can be obtained. After morphologic (opening) operations, the binary images are scanned from top to bottom and filled with "1" between the left and right edges as shown in Figure 2.2.

In the experiment, we set the threshold value as 0.6 for the SSIM maps of both the IR and EO images. The contour of the silhouette is detected with zero-cross based edge detection. The top head points are searched from each frame and used for initial parameter estimation.

2.2.3 Parameter Estimation and Refinement

It is reasonable to assume that the IR and EO cameras are mounted in parallel, which means we can omit the rotation between the frames acquired by the two cameras. Therefore, a 2-D homogeneous transform can describe the geometric relation between the two frames. If IR image is used as a reference, there is:

$$\begin{cases} X_{IR} = kX_{EO} + \Delta X \\ Y_{IR} = kY_{EO} + \Delta Y \end{cases} \tag{2.12}$$

where k stands for the scaling parameter and $\{\Delta X, \Delta Y\}$ are the translating parameters. Thus, there are three parameters to be found in total. The coordinates of the pixels in the IR and EO image are $\{X_{IR}, Y_{IR}\}$ and $\{X_{EO}, Y_{EO}\}$.

Assuming the head top points in IR image correspond to the head top points in EO image, we can solve equation (2.12) with the least square method. Figure 2.3 shows the trajectory of top head points from IR and EO sequences. The initial estimation can be obtained by solving the equation (2.12) given the corresponding head points. However,

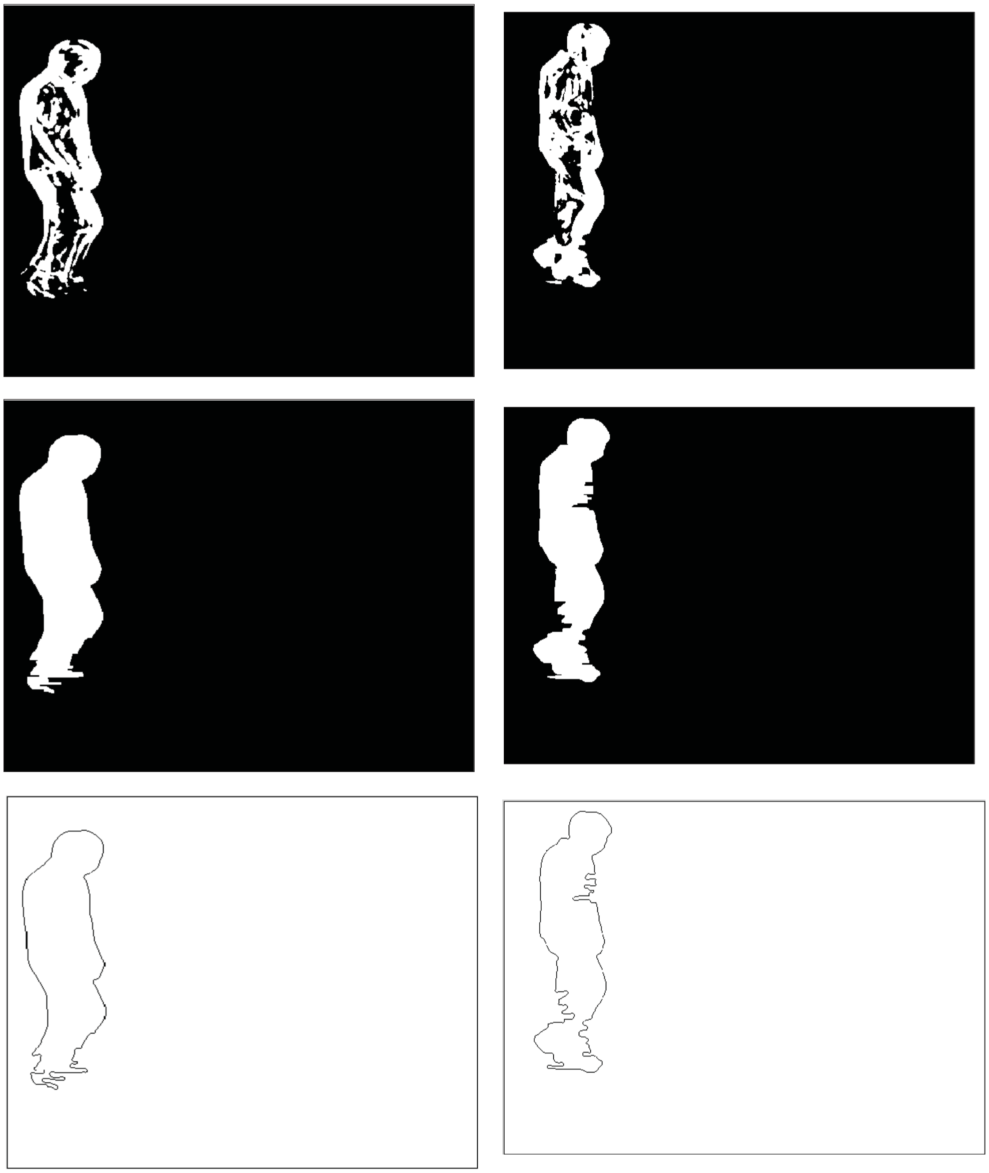
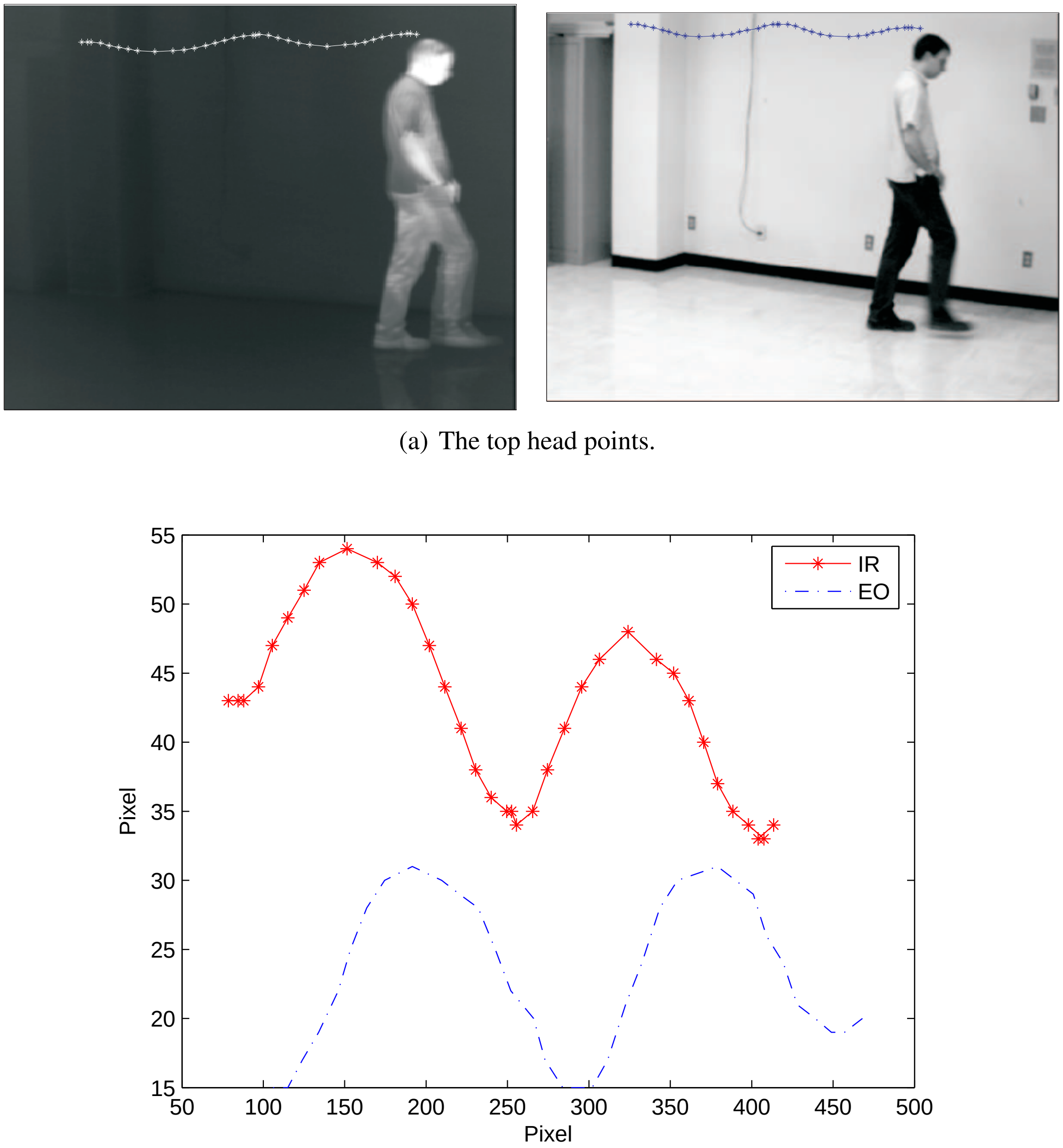


Figure 2.2: The thresolded binary images from SSIM maps are on the top, the processed results on middle, and on bottom are the contours extracted from the processed binary results.



(a) The top head points.

(b) The trajectory of top head points.

Figure 2.3: The top head points in two video sequences.

these points may not be exactly matched. The initial registration can be further refined by applying a mutual information based registration approach [52, 46].

We can use the binary maps in Figure 2.2 to find the region of interest (ROI) from IR and EO images easily as shown in Figure 2.4. Note that binary map can extract the corresponding ROI for any two adjacent frames.

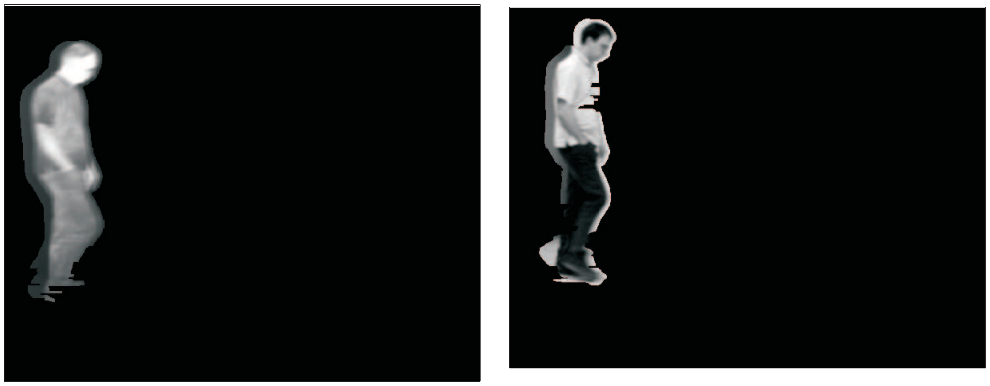


Figure 2.4: The regions of interest from two frames.

The definition of mutual information (MI) for two discrete random variables U and V is:

$$MI\left(U;V\right)=\sum_{v\in V}\sum_{u\in U}p\left(u,v\right)\log_2\frac{p\left(u,v\right)}{p\left(u\right)p\left(v\right)} \tag{2.13}$$

where $p(u,v)$ is the joint probability distribution function of U and V, and $p(u)$ and $p(v)$ are the marginal probability distribution functions of U and V respectively. Actually, MI quantifies the distance between the joint distribution of U and V, i.e. $p(u,v)$, and the joint distribution when U and V are independent, i.e. $p(u)p(v)$. For the IR and EO image, the joint probability distribution can be obtained from the image's histogram. In equation (2.13), $p(u, v)$ can be replaced by the normalized joint grey level histogram of the IR and EO image. There is:

$$p(u, v) \leftarrow h_{IE}\left(l, m\right) = \frac{g\left(l, m\right)}{\sum\limits_{l,m} g(l, m)} \tag{2.14}$$

where $g(l, m)$ is the joint histogram of IR and EO image. Letter l and m correspond to the column and row of a image respectively. The marginal probabilities are represented by normalized marginal histogram of IR and EO image. There are:

$$p(u) \leftarrow h_I(l, m) = \sum_l h_{IE}(l, m) \tag{2.15}$$

$$p(v) \leftarrow h_E(l, m) = \sum_m h_{IE}(l, m) \tag{2.16}$$

Mutual information can be equivalently expressed with joint $\{H(L, M)\}$ and marginal entropies $\{H(L), H(M)\}$ of two variable L and M:

$$MI(L; M) = H(L) + H(M) - H(L, M) \tag{2.17}$$

where there are:

$$\begin{aligned} H(L) &= -\sum_l h_I(l, m) \log_2 h_I(l, m) && (2.18)\\ H(M) &= -\sum_m h_E(l, m) \log_2 h_E(l, m) && (2.19)\\ H(L, M) &= -\sum_{l,m} h_{IE}(l, m) \log_2 h_{IE}(l, m) && (2.20) \end{aligned}$$

The registration is to transform the EO image to the coordinate of the IR image. When the transformed image is aligned with the reference, the MI value is maximized. Thus, searching the transform parameters that maximize MI gives the registration result. Similarly, we use simplex search method as proposed by Chen et al. [52]. The implementation of the simplex search algorithm is available in Matlab® as a function named "fminsearch".

2.3 Experimental Results

The video sequences used in the experiment were captured by the researchers at Laval University. A Radiance PM infrared camera and a Pulnix TMC6700CL camera were used. The Radiance PM camera is a hight resolution, fully calibrated temperature measurement system designed for a wide range of thermal imaging applications. The Pulnix TMC6700CL camera was used to capture EO frames. The sizes of the captured IR and EO frames are 512×460 and 640×480 respectively. The external HSYNC and VSYNC inputs of the Pulnix camera were synchronized on the Radiance NTSC video output signal. The specifications of the two cameras are given in Table 2.1 and 2.2 respectively.

Table 2.1: The configuration parameters for Radiance PM IR camera.

Settings	Radiance PM
sensor	1"
Lens	25 mm
Range	10.0 °C
Level	27.7 °C
AGC	Off
Shutter Speed	1
Palette	Gray

We registered two clips from IR and EO video sequence (30 fps) captured by the two

Table 2.2: The specifications for Pulnix TMC6700CL camera.

Specifications	Pulnix TMC6700CL
Sensor	0.5”
Lens	12 mm
Scanning mode	Progressive
Shutter speed	$1/32 - 1/4000000$

cameras. The threshold value to get the binary image was set as 0.6 for both the clips. The grey level for IR and EO images was rounded to $0 \sim 255$. The initial estimation of the registration parameters from head top trajectory were $\{k = 0.9495; \Delta X = 20.943; \Delta Y = -28.9725\}$. The refinement of this result was carried out for the thirty-five frames in the two clips. The results are shown in Figure 2.5 and the distribution of the parameters is given in Figure 2.6 to 2.8. Table 2.3 lists the mean, maximum, and minimum value of the parameters.

Table 2.3: The registration parameters obtained by maximum MI.

	Mean	Max	Min
k	0.9640	0.9754	0.9552
ΔX	21.19	22.51	16.47
ΔY	-28.25	-25.52	-30.19

The EO frames are transformed and registered with IR frames as shown in Figure 2.9 by using the mean value of the refined registration parameters. The rounding operation is applied to the transformed pixel coordinates. The human body is segmented from IR image and embedded in the EO frames. The synthesized images indicate how well the two sequences are registered. Another method to verify the registration result is to replace the

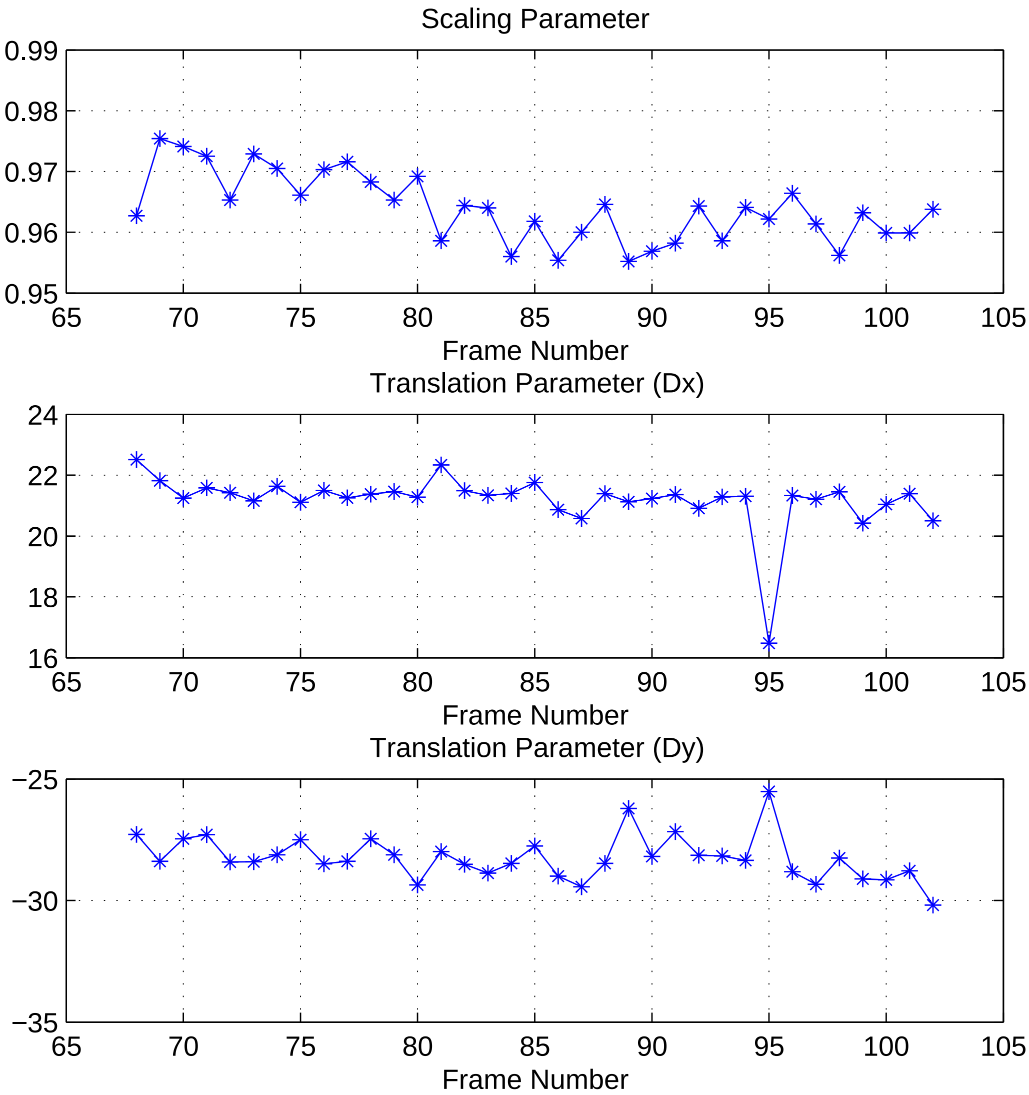


Figure 2.5: The refined registration results based on maximum MI.

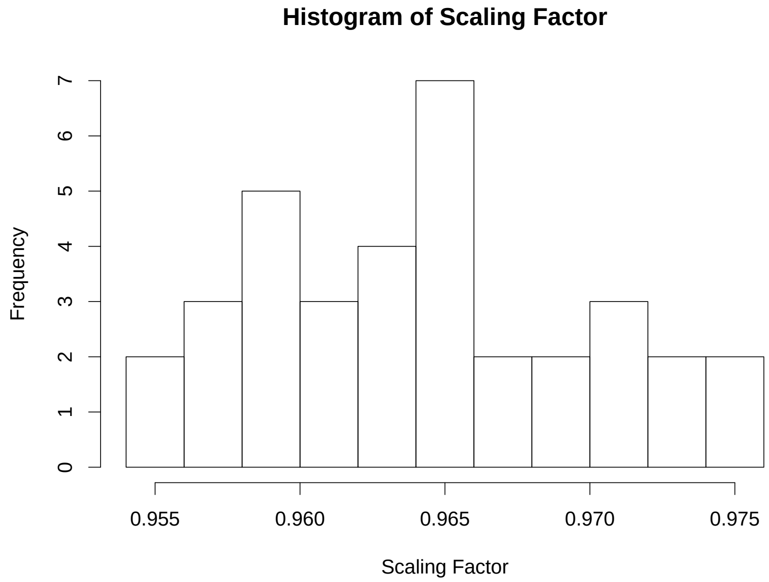


Figure 2.6: The distribution of the refined scaling parameter.

red channel of the visible image with the IR data [47].

2.4 Discussion

The centroid of human body could be another feature point for registration as described in [27]. One precondition is that a “clear” silhouette of human body must be obtained. In the proposed method, the centroid points are not used for registration, because the shadow on the floor makes the bottom boundary indistinct and the centroid point cannot be steadily detected. Obviously, the success of registration depends on how well the feature can be accurately detected. There are a number of factors that affect the registration, for example, the distance from the camera to the moving object.

The translation parameters obtained from 95 frame appears to be a outlier. There are

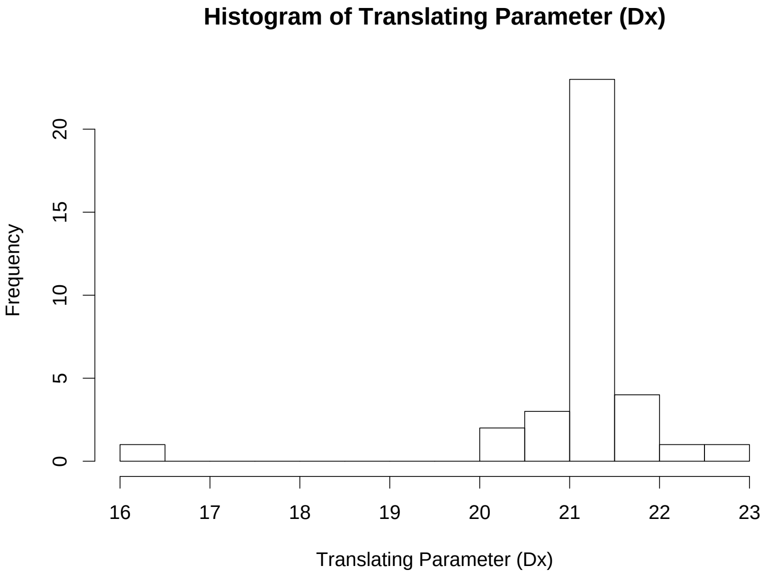


Figure 2.7: The distribution of the refined translating parameter D_x.

two possible reasons contributing to such variance. The first one is that detected head top point may not be accurate. The second is that the shapes of the two silhouette may not be the same.

Although we assume that there is no rotation between two frames, such angular difference may be considered when the registration is refined with maximum MI. In our case, the rotation parameter searched by maximum MI is around 0.0003 rad; therefore, we do not consider it in the experiment.

The registration of multi-modal video sequences does not have to be implemented in real time, only if the configuration of the cameras does not change dynamically. In this chapter, the accuracy of the registration is not studied. It is meaningful to discuss the accuracy when a specific processing is considered. How the accuracy will affect the result

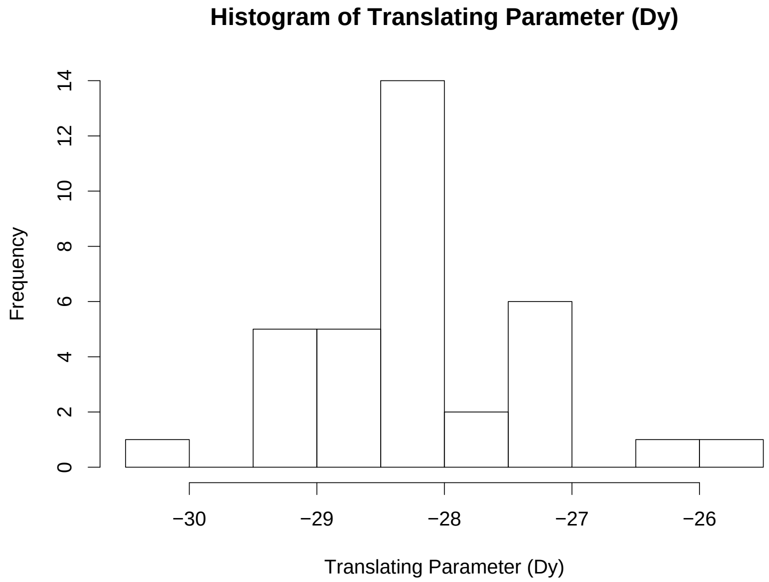


Figure 2.8: The distribution of the refined translating parameter D_y.

of further processing will be investigated in future work. As described in [48], Torresan et al. developed a "master-slave" scheme to alternatively use IR and visual frames for pedestrian detecting and tracking. The visible combination occurs only at blob level; therefore, a coarse (low-accurate) registration should meet the requirement.

2.5 Conclusion

In this chapter, a registration method for multi-sensor video sequences is proposed. The approach is based on registering the trajectories of the head top points detected from the silhouette of frame difference, which is found by the structural similarity measurement. Such differences can be used to find the region of interest. The refinement of the initial

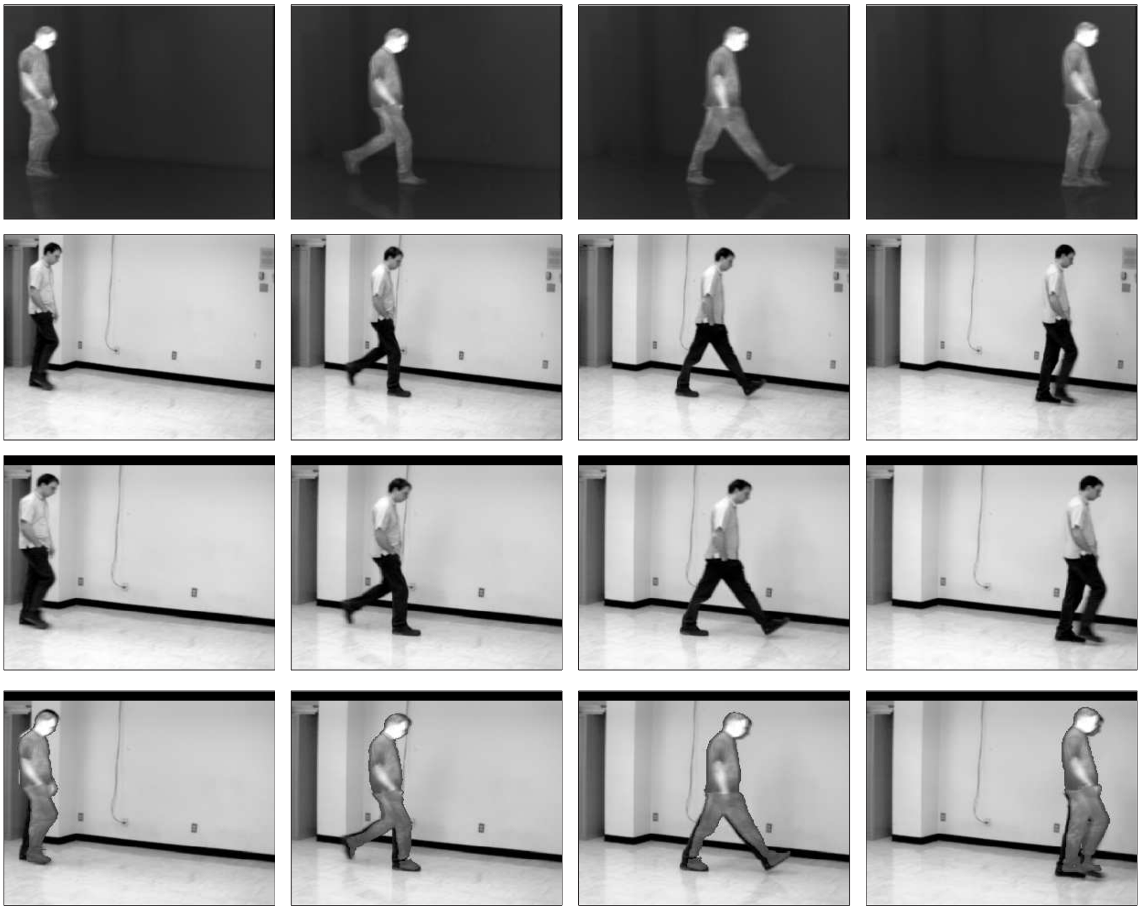


Figure 2.9: The registration results. Top: IR frames; 2^{nd} row: EO frames; 3^{th} row: transformed EO frames; bottom: the synthesized images.

registration is implemented by maximizing the mutual information of the detected regions of interest. The advantage of this technique is that it is not necessary to segment the exact silhouette of the moving object from the video sequence, which is difficult for imaging modality like millimeter wave. Secondly, the proposed method tries to use individual feature point in multiple frames rather than matching multiple points from one image. This makes the registration process easily implemented and the initial searched result is close to the refined one. Although a single feature is used, the registration based on multiple features can be properly implemented. More robust result is expected.

Once the multi-modal images (video sequences) are fully registered; in the next stage, algorithms will be implemented to fuse the input images. Two applications are described in Chapter 3 and Chapter 4 respectively.

Chapter 3

Context Enhancement through Infrared Vision

3.1 Introduction

IN the previous chapter, a solution for registering multi-modal video sequences (images) is proposed, which implements the first step of image fusion and assures the accuracy of pixel-level fusion or other operations. It does not make any sense if the image fusion is implemented without considering any specific requirements from a real application. This chapter deals with the issue of context enhancement for night vision application. The assessment of the performance of different fusion algorithms will be presented in Chapter 6.

The use of thermal vision techniques can be found in numerous applications such as military, law enforcement, surveillance, security, navigation, fire fighting, and wildlife observation [54]. Thermal imaging is a type of infrared (IR) imaging, which detects radiation in the infrared range of the electromagnetic spectrum. The IR image provides an enhanced spectral range that is imperceptible to human beings and contributes to the contrast between objects of high temperature variance and environment. For example, detecting human body in the environment with inadequate illumination is critical for the application like surveillance and intelligent transportation. Once the IR and visual images are fully registered using the method as described in Chapter 2, further process can be carried out.

A well-adopted method to process visual and IR images is the pixel-level fusion [33]. The multi-sensor images must be fully registered as described in Chapter 2. Once the registration process is accomplished, the visual and IR images are then transformed to the wavelet domain. Through combining the coefficients in the transform domain, a new composite image can be obtained by applying the inverse transform. The purpose of the fusion operation is to highlight the objects with high temperature variance and preserve the details of the background. The enhancement may be applied to the visual image before the fusion operation as proposed by Tao et al. [34]. However, there are always questions to the "direct" fusion of visual and IR images at pixel level. In a low thermal contrast environment, the background details like vegetation or soil areas should be represented well in the visual bands [55]. Does the fusion with an IR image contribute to the fidelity of the background objects in such a scenario? The answer might be "no". The fusion may degrade the original information contained in the visual image when they are not complementary. The features presented in the visual band is most suitable for human perception.

Bender et al. conducted a series of tests with a so-called head-tracked vision system, which consisted of thermal and image-intensified TV sensors [56]. The weighted averaging and Laplacian pyramid based fusion implemented by Sarnoff Corporation were tested. Yang and Blum developed a hidden Markov model to correlate the wavelet coefficients across the frequency bands [57]. The expectation-maximization (EM) algorithm was applied to estimate the model parameters and produce the fused image [58]. However, it is hard to tell the difference in the results obtained by different algorithms.

Research on color-based fusion was reported in [59, 60]. The idea is to produce a false color fused image due to the fact that the human visual system is sensitive to colors [60]. Fay and Waxman et al. used the center-surround feedforward shunting network to implement the color composite fusion methods [59]. In [61], Toet et al. demonstrated the improved situational awareness thanks to the image fusion. The details of color-based fusion methods proposed by Toet and Xue can be found in [55] and [60] respectively. In this thesis, we focus on the intensity images.

A modified scheme for the fusion of IR and visual images is proposed in this chapter. The contrast of the visual image is first enhanced through using the pixel value from IR image as an exponential factor. The result is then fused with the visual image again to emphasize the features obtained in the visual band. The objects with high temperature variance are highlighted in the final result. As a result, an enhanced version of the visual image is achieved and can be presented to the end users.

The rest of this chapter is organized as follows. A brief review of the multiresolution image fusion process is given in section (3.2). The feasibility of using adaptive enhancement and image fusion technique for night vision application is investigated in section (3.3). The modified fusion scheme is proposed in section (3.4). Section (3.5) presents more experimental results obtained with proposed method. Discussion and conclusion can be found in section (3.6) and (3.7) respectively.

3.2 Multiresolution Analysis (MRA) based Image Fusion: A Brief Review

The principle for MRA-based fusion methods is to retain the salient image features, which can be easily accessed and manipulated by representing the image in the transform domain. The methods vary with the basis functions and fusion rules. An excellent review of the MRA-based pixel-level fusion by Blum et al. can be found in reference [33]. Piella's overview is another very good reference [62]. The fusion procedure is illustrated in Figure 3.1. The input image $I\left(x,y\right)$ is first represented in the transform domain, i.e. a sum over a collection of functions $g_i\left(x,y\right)$:

$$I\left(x,y\right)=\sum_i y_i g_i\left(x,y\right) \tag{3.1}$$

where y_i are the transform coefficients and can be obtained by projecting the image onto a set of projection functions, $h_i\left(x,y\right)$:

$$y_i=\sum_{x,y} h_i\left(x,y\right) I\left(x,y\right) \tag{3.2}$$

The fusion rule is then applied to y_i based on the measurement of image features and characteristics of $g_i\left(x,y\right)$. After applying the inverse transform, the fused image is obtained.

For pixel-level fusion, the outcome of the fusion process is also an image, which should be more suitable for further analysis than any input. A comparison of some MRA fusion algorithms is summarized in Table 3.1 and 3.2 respectively. Generally, the study of MRIF is twofold, encompassing a multiresolution algorithm and a coefficient combination rule. A number of MRA algorithms have been investigated for the fusion of multi-sensor images

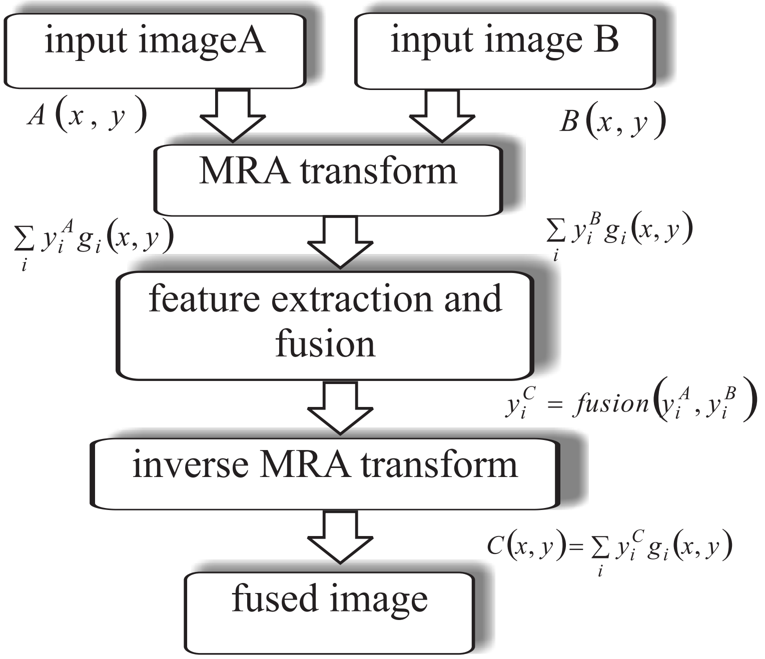


Figure 3.1: The procedure of MRA-based pixel level fusion.

so far. For detailed implementation, relevant references are listed in Table 3.1 and 3.2. The choice of the MRA algorithms largely depends on the characteristics of the algorithm and the signal to be processed. Since an image is represented as a weighted sum of basis functions, choosing the basis function that resembles the signal will facilitate the analysis. The major steps of MRIF include: image decomposition, coefficient combination, and image reconstruction. The basic rule for coefficient combination is the absolute value maximum selection in the high-frequency (high- and band-pass) bands and averaging in the low-frequency (low-pass) band, i.e. the coefficients with larger absolute value from the high frequency bands will be retained and used for reconstruction, because the larger values correspond to image features like edges, lines, or boundaries. More sophisticated rules will consider the area or region around the pixel and the corresponding areas or regions across the frequency bands or resolution scales [63, 62]. As a result, image feature measurements in a region or across the frequency bands are generated. A selection rule is created or the weighting coefficients are derived from such measurements.

Table 3.1: Comparison of multiresolution image fusion schemes: image pyramid.

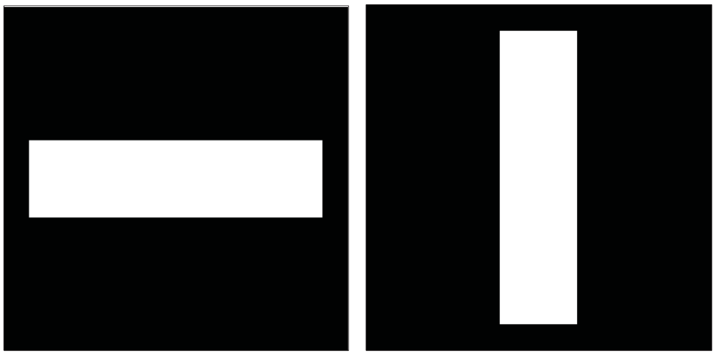
MRA	Fusion rule	Contribution	Evaluation	Applications
Laplacian pyramid [64]	absolute value maximum selection (AVMS)	first study on MRIF		multi-focused images
Ratio-of-lowpass pyramid [65, 66]	maximum absolute contrast selection	use of RoLP	subjective evaluation	simulation
Gradient pyramid	fusion based on match and salience measure [5]	image feature based fusion		fusion of IR and visible image, multi-exposure, multi-focus images
	weighted average [67, 68]	perceptual-based fusion	SNR	hyperspectral image
Morphological pyramid [13]	maximum operation	use of morphological pyramid	cross-correlation	CT and MRI images
Steerable pyramid [69]	apply Laplacian pyramid and AVMS rule for sub-images	iterative fusion of sub-images	RMSE	standard images for simulation, multi-sensor images

To illustrate the fusion process, a simple example is given below in Figure 3.2. Two images with a horizontal and vertical square bar cross the center are fused with six MRA-based fusion algorithms. The six algorithms include Laplacian pyramid (LAP), gradient-based pyramid (GRAD), ratio-of-lowpass pyramid (RoLP), Daubechies wavelet four (DB), shift-invariant discrete wavelet transform (SIDWT), and steerable pyramid (STEER) [64, 5, 70, 73, 69]. The decomposition level is selected as four and the fusion rule is to select

Table 3.2: Comparison of multiresolution image fusion schemes: discrete wavelet.

MRA	Fusion rule	Contribution	Evaluation	Applications
Orthogonal wavelet [70]	AVMS	consistency verification, concept of region-based fusion	RMSE	multi-focus images, multi-sensor images
Steerable dyadic wavelet [71, 14]	maximum local oriented energy	image feature represented with oriented energy	MSE	different channels of landsat TM images
Discrete wavelet frame	Rockinger [72, 73, 74]; Fusion rule is the same as [5].	studies on temporal stability and consistency	image sequences	
	activity measure, region-based rule, grouping approach [75]	studies on region-based approach and grouping method	RMSE, mutual information, percentage of correct decision	multi-focus images, millimeter wave images, infrared images
Contrast-based wavelet [76]	absolute value maximum selection	present the concept of directive contrast	SNR	infrared and visual images
Complex wavelet [77]	chain representation fusion	use of complex wavelet		multi-focus images, CT and MR images
Multiwavelet [78]	pixel selection based on the image's feature map	use of multiwavelet	subjective evaluation	SPOT images

the absolute maximum for the high- and band-pass sub-images (components) and average the low-pass sub-images (components).

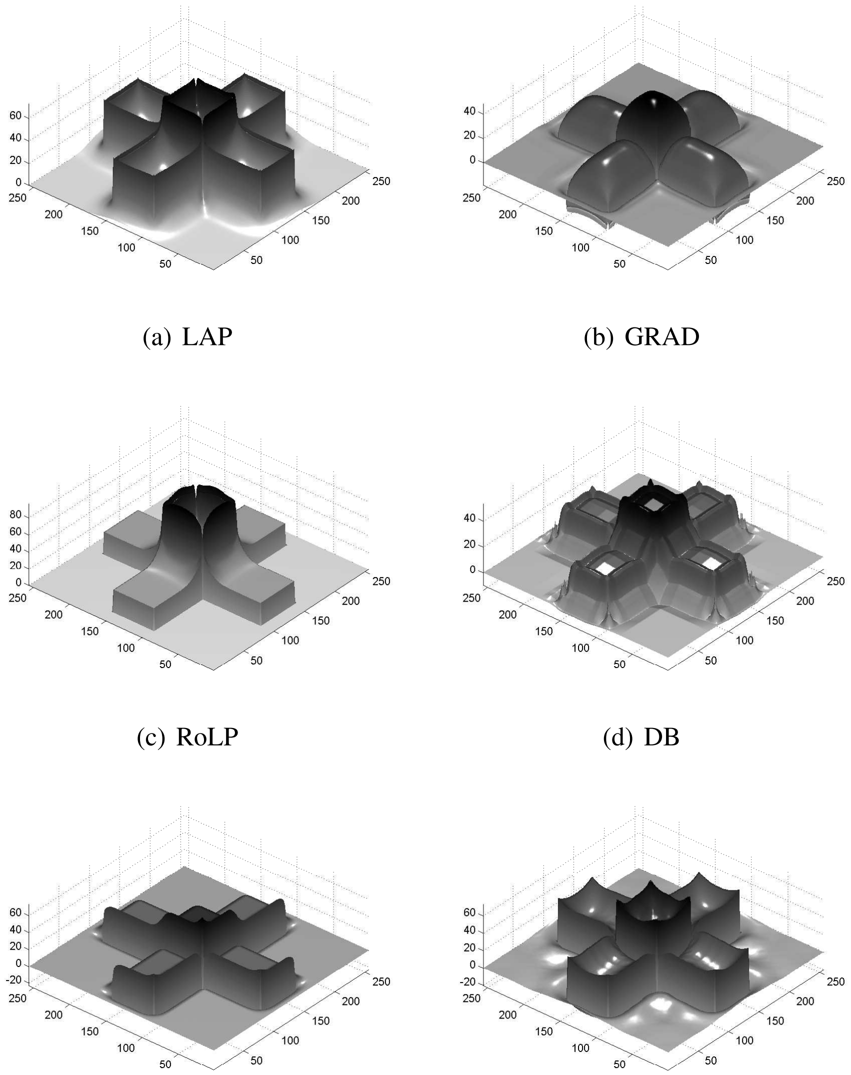


(a) Horizontal bar. (b) Vertical bar.

Figure 3.2: Two images are used for testing MRA-based image fusion.

To clearly show the results, we visualize the fused images in three dimensions in Figure 3.3. For this application, Figure 3.3(a) and 3.3(f) give a better results, which present a shaper edge for the two blocks. The fused result depends on how the features are represented by the MRA algorithms. In other words, the same features are treated differently in different MRA algorithms even though the same fusion rule is applied. To demonstrate how the fusion rule affects the fusion result, the maximum selection of all the coefficients is implemented for the steerable pyramid based fusion as shown in Figure 3.4. This is not a benchmark study for the MRA-based fusion. The simple example demonstrates how the fusion algorithms work and what can be achieved eventually. Readers are referred to the references for detailed implementation and discussion. Rockinger's Matlab® toolbox is a good reference for practice as well [79].

The concept of match measure and salience measure originated from Burt's work on gradient pyramid based image fusion [5], where the match measure determined the selection or averaging operation while the salience measure chose the coefficients for the



(a) LAP

(b) GRAD

(c) RoLP

(d) DB

(e) SIDWT

(f) STEER

Figure 3.3: The fusion results with different MRA-based fusion algorithms.

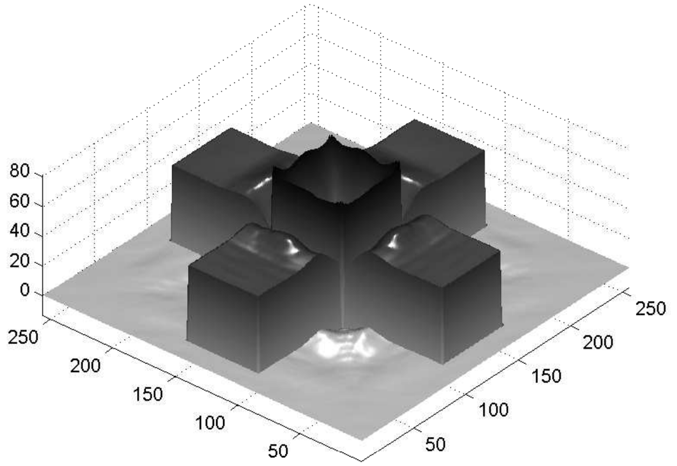


Figure 3.4: The fusion result with the steerable pyramid. The fusion rule is the maximum selection of both the low-pass and high-pass coefficients (see Figure 3.3(f) for comparison).

reconstruction in the selection mode. Wilson et al. introduced the contrast sensitivity measure to weight the coefficient sets [67, 68]. Li's rule for coefficient selection was based on a 3-by-3 or 5-by-5 window [70], where the pixel with the maximum absolute value in the window represented the activity of the pixel located at the center. Li also introduced a consistency verification as a rectification of the selection process. Zhang and Blum used the average value in the region contoured and segmented by the Canny edge detector instead of the maximum pixel value [75]. Thus, the approach is more robust to the noise. Koren used the local oriented energy as a metric of image feature and the coefficient selection was based on such measurement [71]. Cross-band selection and coefficient grouping methods were proposed by Xydeas and Zhang respectively [80, 75]. This is actually another consideration for the region effect, since a single pixel at a lower resolution corresponds to several pixels (region) at a higher resolution. Yang and Blum recently proposed to use the hidden Markov model (HMM) to capture the correlation of the coefficients across the resolution scales [57]. Although different MRA algorithms are still being proposed in various publications, no benchmark study has been carried out so far.

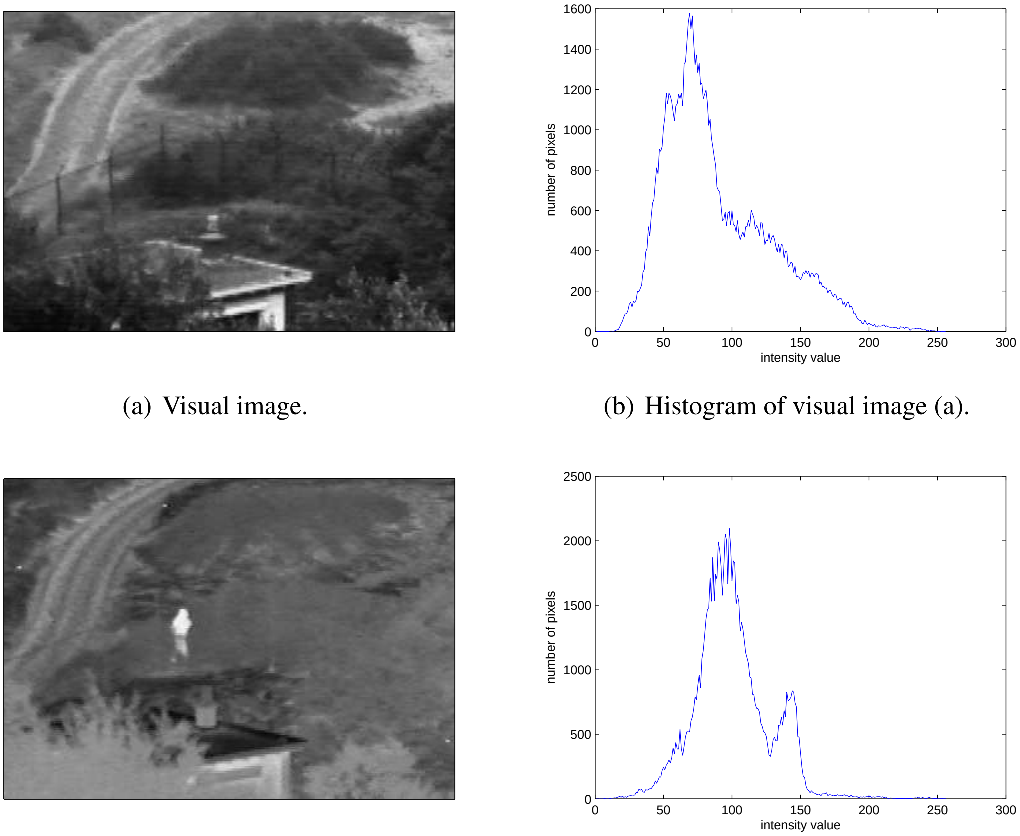
Recall the two types of image fusion described in Chapter 1, the MRIF implements the first type of fusion, i.e. feature combination or integration. Through comparison, salient features are retained in the fused result. However, the salience does not necessarily mean useful. For example, if the noise is salient, it will still be kept in the result as well. This is a limitation of the MRIF method.

3.3 Enhancement and Fusion

3.3.1 Histogram-based Operations

Histogram-based operations like histogram equalization and matching provide a basic tool for image enhancement. The description of such operations is available in most of the text books on image processing, for example, reference [81], and will not be repeated here again. We use two images from two video sequences to demonstrate the processing methods [79]. One is a visual image as shown in Figure 3.5(a) and the other one in Figure 3.5(c) is from an infrared camera. Due to the inadequate illumination, the human body appearing in the IR image cannot be identified from the visual image, but is visible in the IR image. In Figure 3.5, the histograms of the two images are plotted.

First, the histogram equalization is applied to the visual image. The equalization transforming of the intensity values is to match a flat histogram with 64 bins. The result is presented in Figure 3.6(a) and the corresponding histogram can be found in Figure 3.6(b). However, the human body is still hard to identify. The histogram of the visual image is then



(a) Visual image.

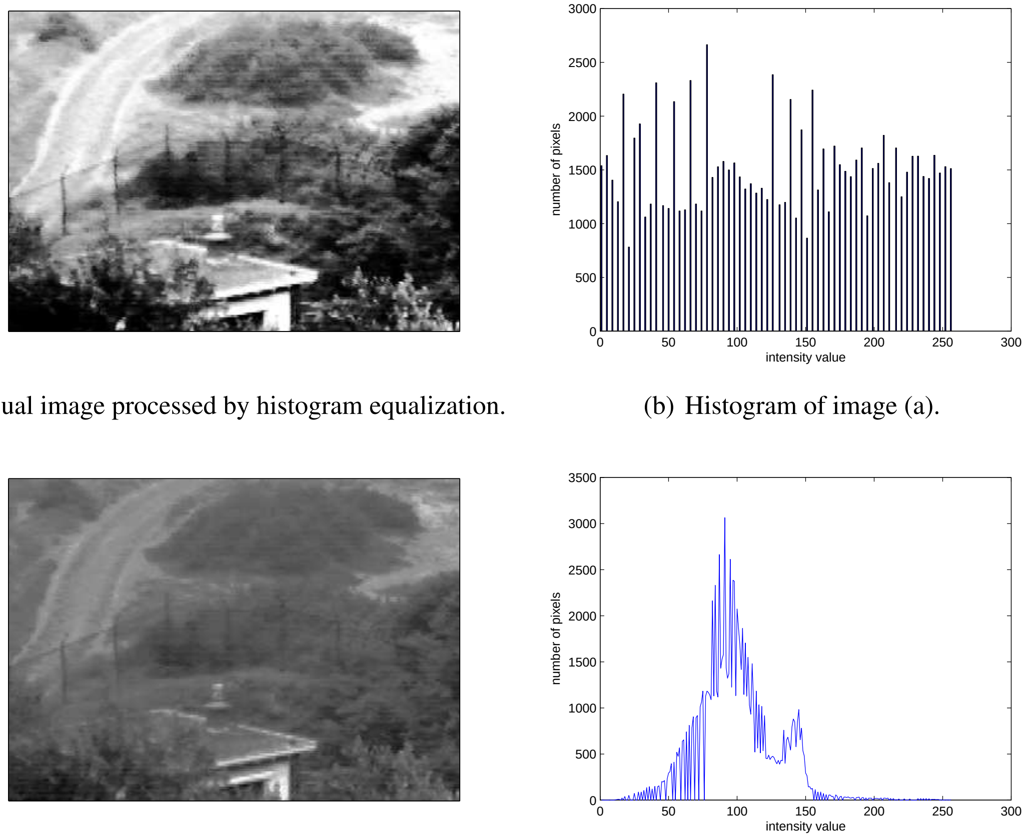
(b) Histogram of visual image (a).

(c) Infrared image.

(d) Histogram of infrared image (c).

Figure 3.5: The visual image and infrared image.

manipulated to match the histogram of the IR image. The results are given in Figure 3.6(c) and 3.6(d) respectively. Similarly, no significant improvement is achieved.



(a) Visual image processed by histogram equalization.

(b) Histogram of image (a).

(c) Visual image processed by histogram matching.

(d) Histogram of image (c).

Figure 3.6: The histogram-based processing of visual image.

3.3.2 Adaptive Enhancement

As the real-world scenes exhibit with high dynamic range radiance spanning more than six orders of magnitude, the processing of the image can be implemented through compression of such dynamic range [82]. Herein, we tested two methods proposed by Tao et al. in [82, 83, 34, 84]. The two approaches are the "adaptive and integrated neighborhood dependent approach for nonlinear enhancement" (AINDANE) and an nonlinear enhancement based on an illuminance-reflectance model. The basic ideas of the two approaches are quite similar, i.e. applying a nonlinear transfer function to compress the dynamic range. The implementation procedures of the two algorithms are illustrated with the flowcharts shown in Figure 3.7. The major difference between these two approaches is the nonlinear function, which is highlighted with a dashed square. For an image $I(x, y)$, the final enhancement is implemented with the following equation [82]:

$$S\left(x,y\right) = 255 I_n'\left(x,y\right)^{E(x,y)} \tag{3.3}$$

where $S\left(x,y\right)$ is the enhanced image and $E\left(x,y\right)$ is obtained by the following two equations:

$$E\left(x,y\right) = r\left(x,y\right)^P = \left[\frac{I_G\left(x,y\right)}{I\left(x,y\right)}\right]^P \tag{3.4}$$

$$I_G\left(x,y\right) = G\left(m,n\right) * I\left(x,y\right) \tag{3.5}$$

Herein, $G\left(m,n\right)$ is a $m \times n$ Gaussian kernel and parameter P is an empirical parameter [82]. The original image is represented as $I\left(x,y\right)$. In the ANIDANE algorithm, parameter P is given by:

$$P = \begin{cases} 3 & \sigma \leq 3 \\ \frac{27-2\sigma}{7} & 3 < \sigma < 10 \\ 1 & \sigma \geq 10 \end{cases} \tag{3.6}$$

where σ is the global standard deviation of the image. Image $I_n\left(x,y\right)$ is obtained by normalizing $I\left(x,y\right)$ to the range of $[0,1]$. The nonlinear transfer function is given by:

$$I_n'\left(x,y\right)=f\left(I_n\left(x,y\right),z\right)=\frac{I_n\left(x,y\right)^{\left(0.75z+0.25\right)}+\left(1-I_n\left(x,y\right)\right)0.4\left(1-z\right)+I_n\left(x,y\right)^{\left(2-z\right)}}{2} \tag{3.7}$$

where the parameter z is determined by:

$$z=\begin{cases} 0 & L\leq 50 \\ \frac{L-50}{100} & 50<L\leq 150 \\ 1 & L>150 \end{cases} \tag{3.8}$$

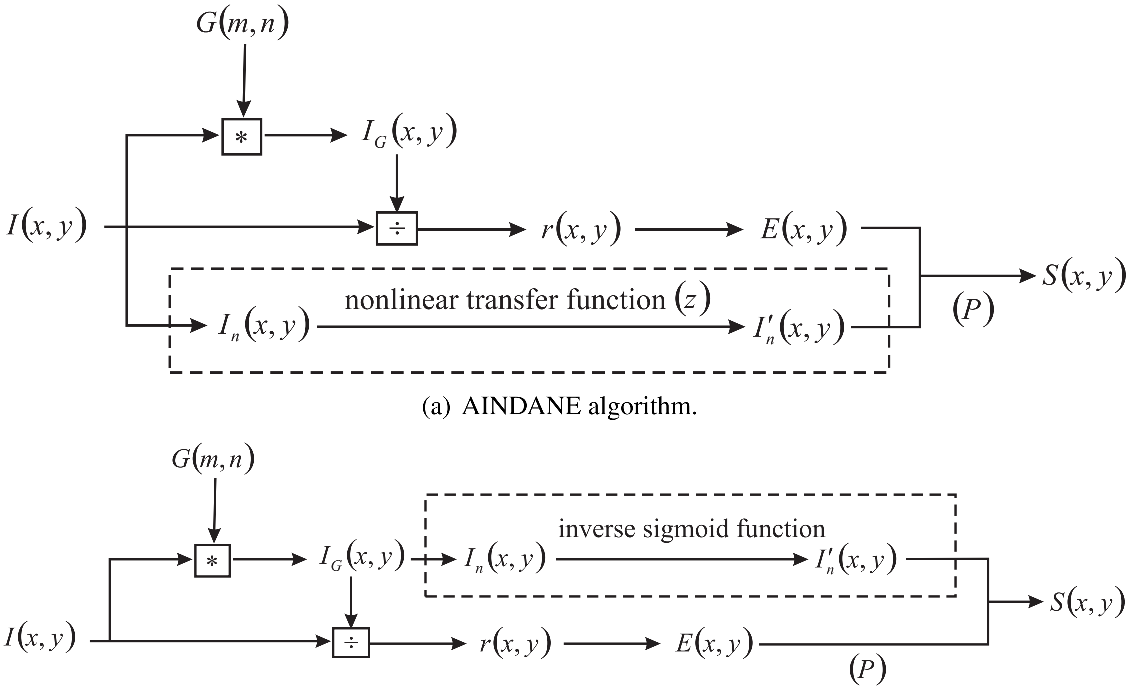
Herein, the intensity level L corresponds a value of 0.1 in the cumulative distribution function of the image. In the method using illuminance-reflectance model, the P value becomes [83]:

$$P=\begin{cases} 2 & \sigma\leq 30 \\ -0.03\sigma+2.9 & 30<\sigma\leq 80 \\ 1/2 & \sigma>80 \end{cases} \tag{3.9}$$

An inverse sigmoid function is used to obtain $I_n'\left(x,y\right)$. This function can be expressed with the following equation:

$$I_n'\left(x,y\right)=\frac{-\frac{1}{a}\ln\left[\frac{1}{I_n\left(\frac{1}{1+e^{-av_{max}}}-\frac{1}{1+e^{-av_{min}}}\right)+\frac{1}{1+e^{-av_{min}}}}-1\right]-v_{min}}{v_{max}-v_{min}} \tag{3.10}$$

The parameter v_{max}, v_{min}, and a can be tuned manually. In the experiment, v_{max} and a were selected as 3 and 1 respectively. The value of v_{min} is determined by the global mean of the image $I\left(x,y\right)$, i.e. I_m, as:



(a) AINDANE algorithm.

(b) Enhancement based on an illuminance-reflectance model.

Figure 3.7: The adaptive image enhancement algorithms.

$$v_{\min} = \begin{cases} -6 & I_m \leq 70 \\ \frac{I_m - 70}{80} \times 3 - 6 & 70 < I_m < 150 \\ -3 & I_m \geq 150 \end{cases} \tag{3.11}$$

Figure 3.8 shows the results achieved by applying the two adaptive enhancement approaches. These images do not show the details of the poorly illuminated regions, although they are enhanced to some extent. To achieve an optimal enhancement, users need to manually adjust the parameters used in the adaptive enhancement algorithms and this may vary with images. Unfortunately, the sensitivity of these parameters is not reported by the authors.

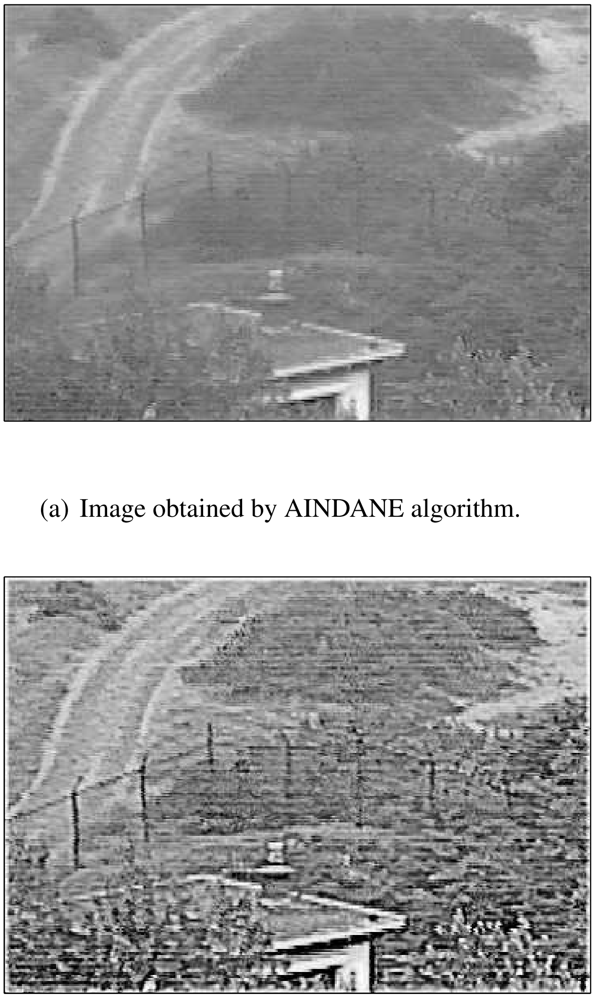
3.3.3 Pixel-level Image Fusion

Multiresolution pixel-level fusion is often employed to combine visual and IR images. Although there are a number of multiresolution fusion algorithms available, a benchmark study of those fusion methods is not available and beyond the scope of this chapter. In this work, the fusion result of the visual and IR image achieved by the steerable pyramid [69] is presented.

The major problem of orthogonal wavelet is the lack of shift invariance, i.e. the translation of the input signal does not correspond to the translation of the output signal. To overcome such limitations, Simoncelli and Freeman et al. proposed a so-called steerable pyramid [85, 86]. The basis functions of the steerable pyramid are directional derivative operators of different sizes and orientations. The steerable pyramid performs a polar-separable decomposition in the frequency domain, thus allowing independent representation of scale and orientation [87]. Such representation is invariant with respect to translation and rotation. An image of N level decomposition can be represented as:

$$I\left(x,y\right)\rightarrow\left(\;LI\left(x,y\right),\;\;BI_{i}^{j}\left(x,y\right)\Big|_{i=1\ldots N}^{j=1\ldots K}\;,\;\;HI\left(x,y\right)\;\right) \tag{3.12}$$

This overcomplete representation consists of three parts: one low-pass component $LI(x,y)$, one high-pass component $HI(x,y)$, and $K\times N$ band-pass components $\left\{BI_{i}^{j}\left(x,y\right)\Big|_{i=1\ldots N}^{j=1\ldots K}\right\}$. For each level, one band-pass component corresponds to an orientation angle $\left(i-1\right)\pi/4$, where $i=1\ldots K$. Oriented features can be extracted by using the steerable pyramid representation. The structure of the steerable pyramid is shown in Figure 3.9, where one high-pass filter $H_0\left(\omega\right)$, two low-pass filters $L_0\left(\omega\right)$ and $L_1\left(\omega\right)$, and K band-pass filters $B_k\left(\omega\right)$



(a) Image obtained by AINDANE algorithm.

(b) Image enhanced based on an illuminance-reflectance model.

Figure 3.8: The enhancement results of visual image achieved by adaptive algorithms.

are involved.

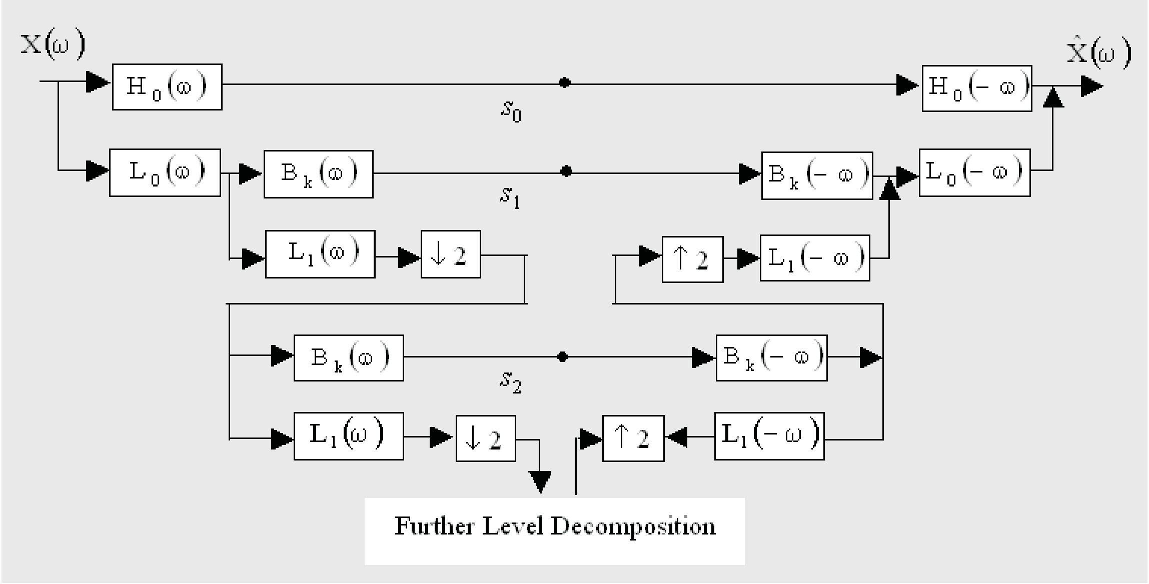


Figure 3.9: The architecture of the steerable pyramid.

To eliminate aliasing, avoid amplitude distortion, and cascade the system recursively, the following constrains should be satisfied [86, 88]:

$$\L_1(\omega) = 0\; for\ |\omega| > \frac{\pi}{2} \tag{3.13}$$

$$|H_0(\omega)|^2 + |L_0(\omega)|^2 = 1 \tag{3.14}$$

$$|L_1(\omega)|^2 + \sum\nolimits_{k=1}^{K} |B_k(\omega)|^2 = 1 \tag{3.15}$$

References [86, 85] give more information about the steering theory and details of filter design. An example of four-level decomposition is given in Figure 3.10. In this case, four oriented filters are employed. The low-pass component is also shown. The algorithm can be efficiently implemented with FPGA (field-programmable gate array) as proposed in [89]. In this study, we use Matlab® to test our algorithms.

The fusion rule consists of the absolute maximum value selection (AMVS) for the high-pass sub-bands ($y_{i,high}(x,y)$) and the average for the low-pass band ($y_{low}(x,y)$). Written in mathematic formulas, there are:

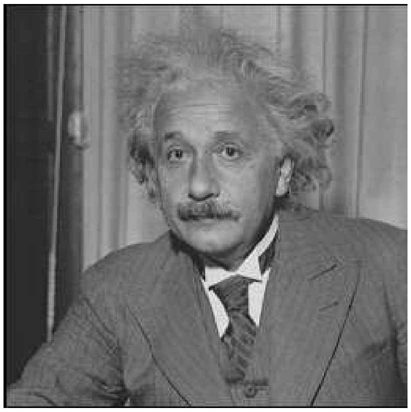
$$y_{i,high}^{new}\left(x,y\right)=\begin{cases} y_{i,high}^{A}\left(x,y\right), & F_i^A\left(x,y\right)\geq F_i^B\left(x,y\right) \\ y_{i,high}^{B}\left(x,y\right), & Others \end{cases} \tag{3.16}$$

$$y_{low}^{new}\left(x,y\right)=\frac{1}{2}\left(y_{low}^{A}(x,y)+y_{low}^{B}(x,y)\right) \tag{3.17}$$

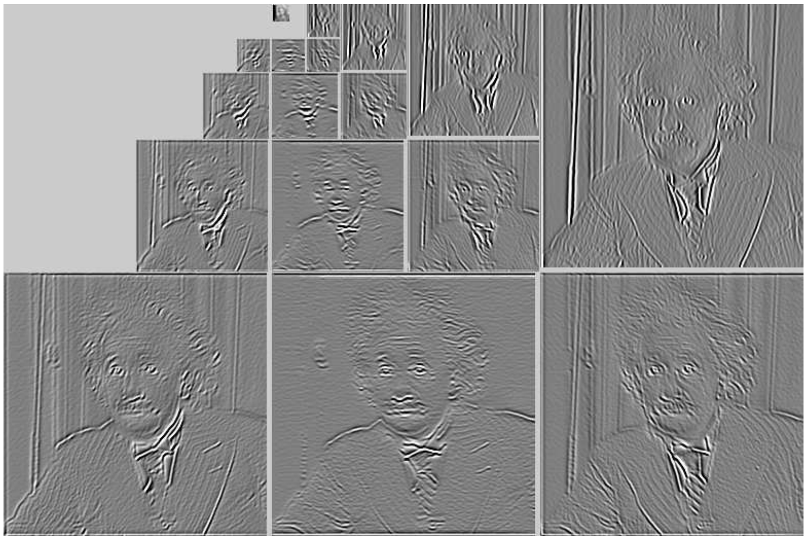
In this case, $F_i(x,y)$ is the absolute value of $y_{i,high}(x,y)$ and i refers to the scale level of the decomposition. The pixel-level fusion result is given in Figure 3.11. Although the human body is perceptible in the fused image, this image does not present all the features appearing in the visual band. This image is not the scene that we see with our eyes.

3.4 A Modified Scheme

In the multiresolution fusion, the averaging of low-pass sub-images incorporates the contrast of the IR image while the AMVS scheme retains the details from both the IR and visual images. There is also question to such operation. If the purpose is only to find the object of high temperature variance, the IR image should be enough and a visual image is not needed. However, the IR image is not what people perceive with their eyes and a visual image is most appropriate for human perception. The details from IR do destroy the contents of the fused image. The advantage of directly fusing IR and visual image at pixel

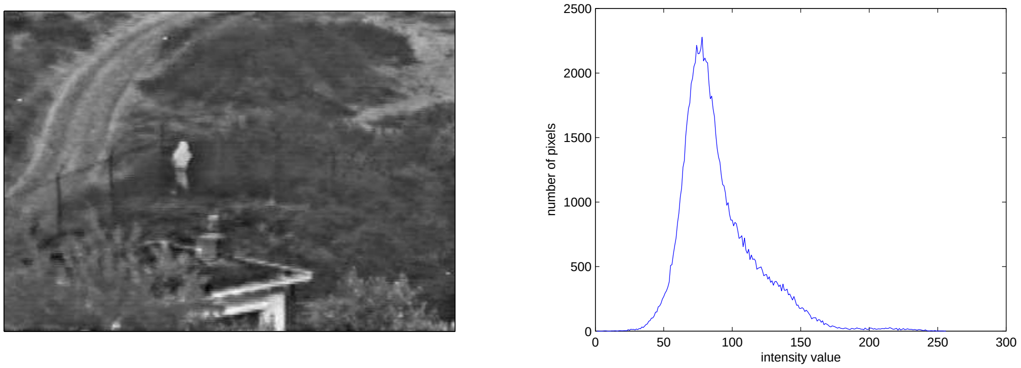


(a) The original image: Einstein.



(b) The decomposition.

Figure 3.10: An example of steerable pyramid decomposition.



(a) The fused image.

(b) Histogram of image (a).

Figure 3.11: The pixel-level fusion of visual and IR images.

level is ambiguous.

Therefore, a modified fusion method is proposed to overcome this problem. The visual image is first enhanced by incorporating the information from IR image. To retain the details from the visual image, the enhanced image is fused with the visual image again by applying the multiresolution fusion algorithm. This method implements the enhancement and fusion, but it is totally different from the approach in [34], where visual image is enhanced first and then fused with the IR image. The approach presented in [34] meets the same problem as the direct fusion.

To enhance the visual image, the visual image, $I_{visual}(x, y)$, and IR image, $I_{IR}(x, y)$, are first normalized to the range $[0, 1]$ and we get $I'_{visual}\,(x, y)$ and $I'_{IR}\,(x, y)$. Motivated by the adaptive enhancement strategy as shown in equation (3.3), we used the normalized IR image as the exponential factor. Then, the enhanced image is obtained from:

$$I_{visual}^{en}\left(x,y\right)=255I_{visual}'\left(x,y\right)^{I_{IR}'(x,y)} \tag{3.18}$$

For a wide-area surveillance application, detection of human body, which is of higher temperature variance, is achieved by identifying the pixels with larger intensity value in the IR image. The exponential function can enhance this difference in the visual image.

Such enhancement can be illustrated with Figure 3.12. If we have two pixel values from IR and visual image, the enhanced pixel value can be easily located from the mesh surface in Figure 3.13.

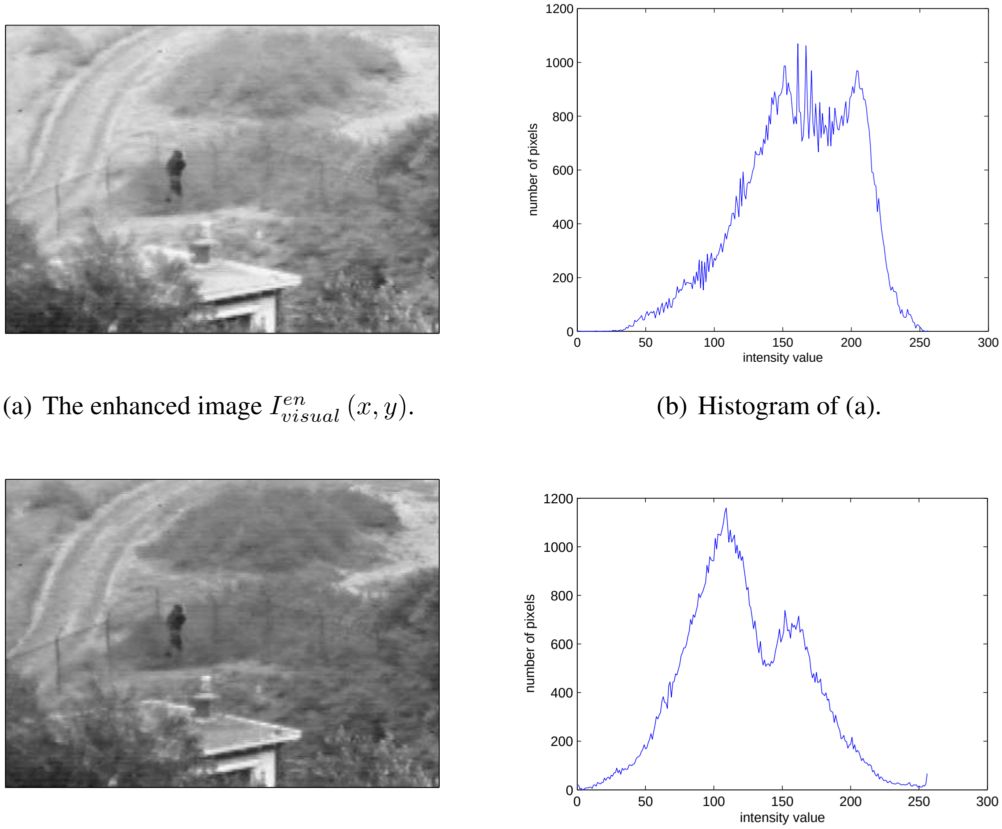
Then, the enhanced image $I_{visual}^{en}\left(x,y\right)$ is fused with the visual image $I_{visual}\left(x,y\right)$ by using the steerable pyramid based fusion algorithm (*steer_fuse*) [69]:

$$S\left(x,y\right)=steer\_fuse\left(I_{visual}\left(x,y\right),I_{visual}^{en}\left(x,y\right)\right) \tag{3.19}$$

In the implementation, four oriented band-pass filters are employed in the steerable pyramid algorithm. The decomposition level is four. The same setup is used throughout the experiment. The final result is shown in Figure 3.12(c). The background information is retained well while the human body is successfully highlighted.

3.5 More Results

The modified scheme is validated with nine groups of images (Figure 3.14-Figure 3.22). For each group there are four images: (a) the visual image; (b) the infrared image; (c) the result of MRA-based pixel-level fusion result; and (d) obtained with the proposed method. We assume the visual and the infrared images in the experiments are perfectly registered. The nine groups of images are retrieved from [79].



(a) The enhanced image $I_{visual}^{en}(x, y)$.

(b) Histogram of (a).

(c) The fused image $S(x, y)$.

(d) Histogram of (c).

Figure 3.12: The result achieved by modified fusion method.

Compared to the MRA-based pixel-level fusion, the proposed method presents the results in the visual spectrum band, which is more suitable for human perception. Hidden features in the visual image are highlighted in the fused result. The only exception is the example shown in Figure 3.17. The human body is not perceptible in the fused results obtained by either of the two methods. This will be discussed in next section.

Another observation is that the MRA-based pixel-level fusion mixed the two intensity

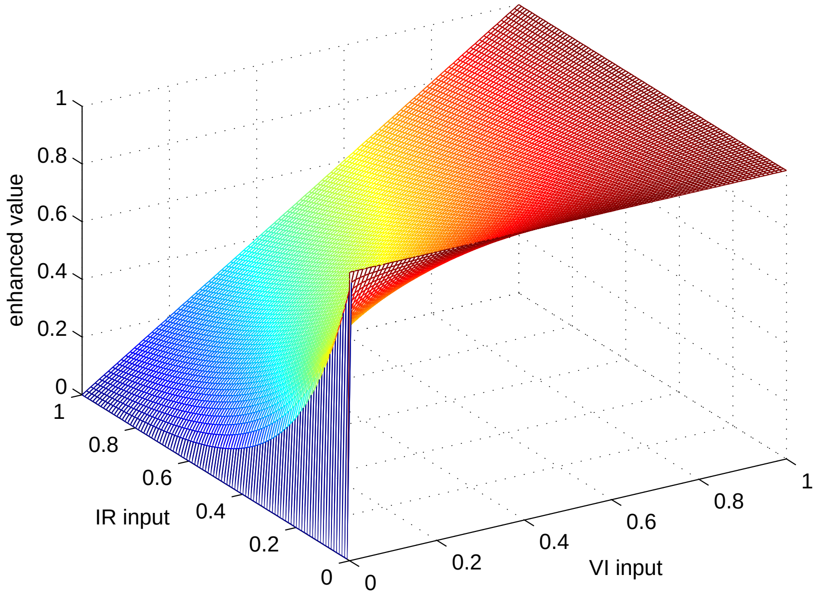
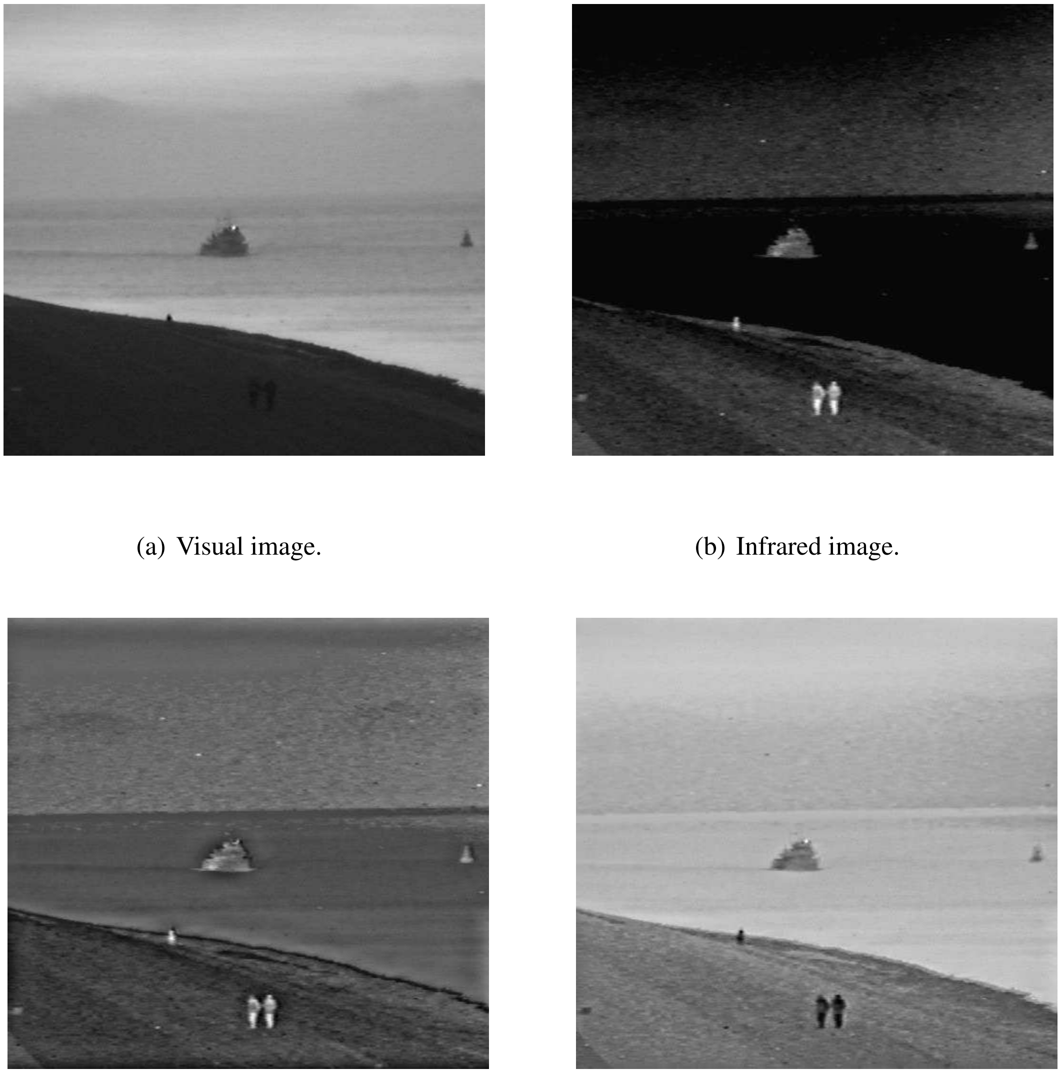


Figure 3.13: The enhancement function.

tables of the visual and IR image. The human body in the fused image exhibits a quite similar appearance to that in the IR image. In contrast, our proposed method achieves a modification of the pixel value in the visual image. The details depend on the visual image only.

3.6 Discussion

In the results shown in Figure 3.17, neither the MRA-based pixel-level fusion nor the modified method can identify the human body, which is detectable in the IR image. In the



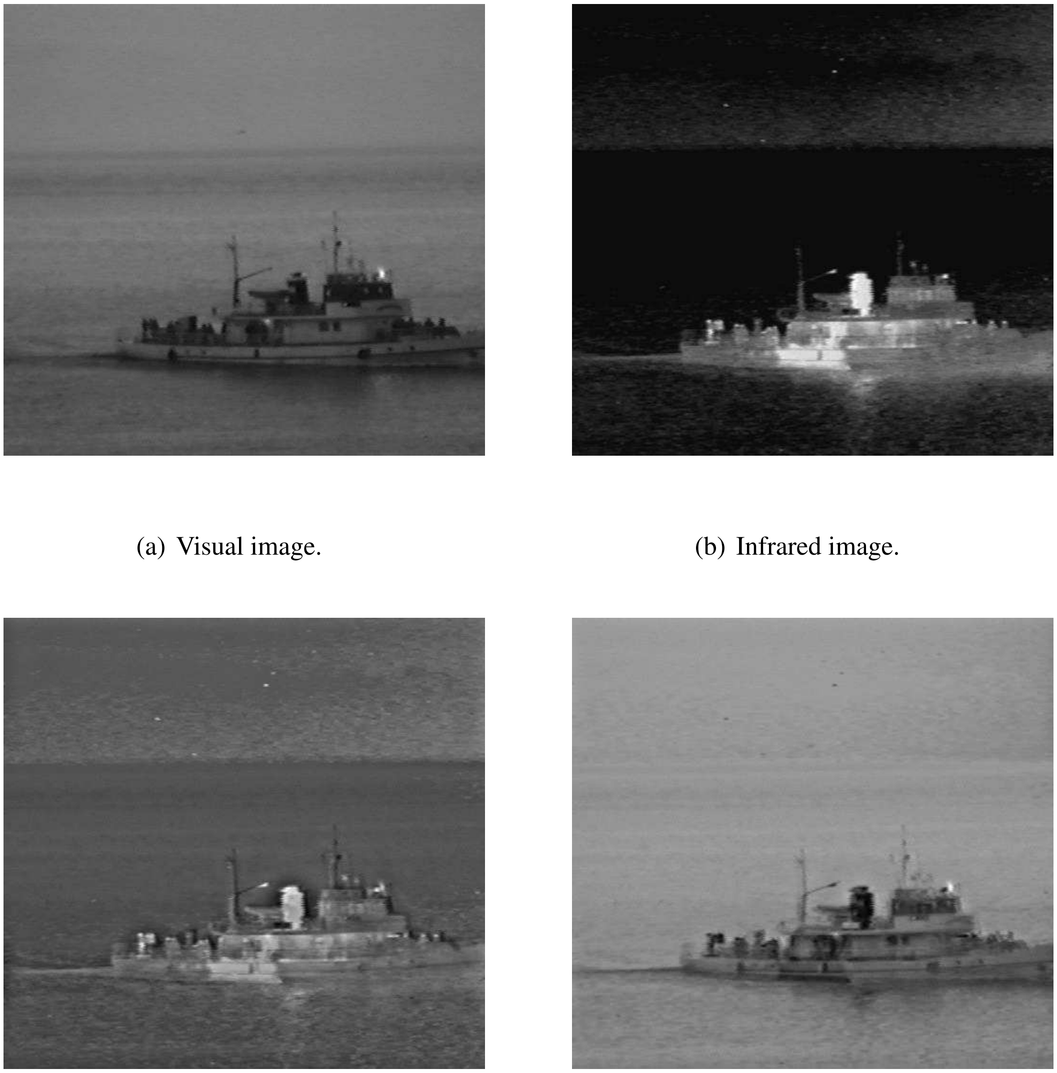
(a) Visual image.

(b) Infrared image.

(c) MRA fusion result.

(d) Modified fusion result.

Figure 3.14: TNO Kayak (frame 7118a).



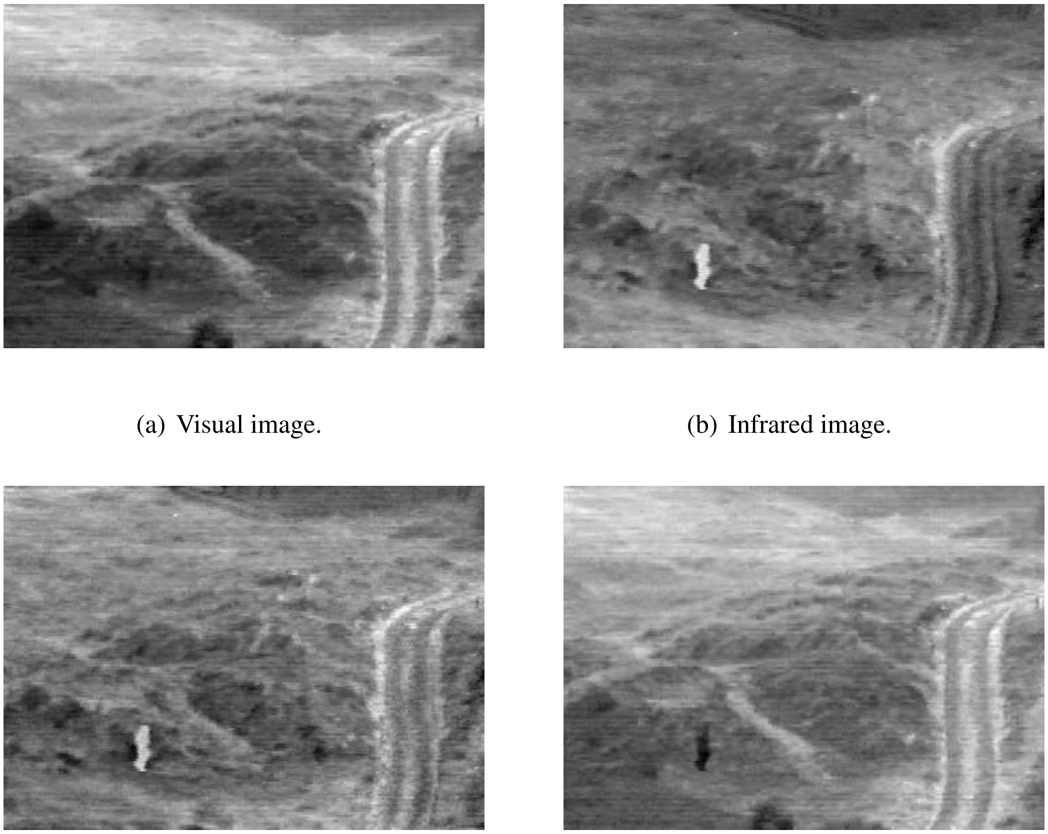
(a) Visual image.

(b) Infrared image.

(c) MRA fusion result.

(d) Modified fusion result.

Figure 3.15: TNO Kayak (frame 7436a).



(a) Visual image.

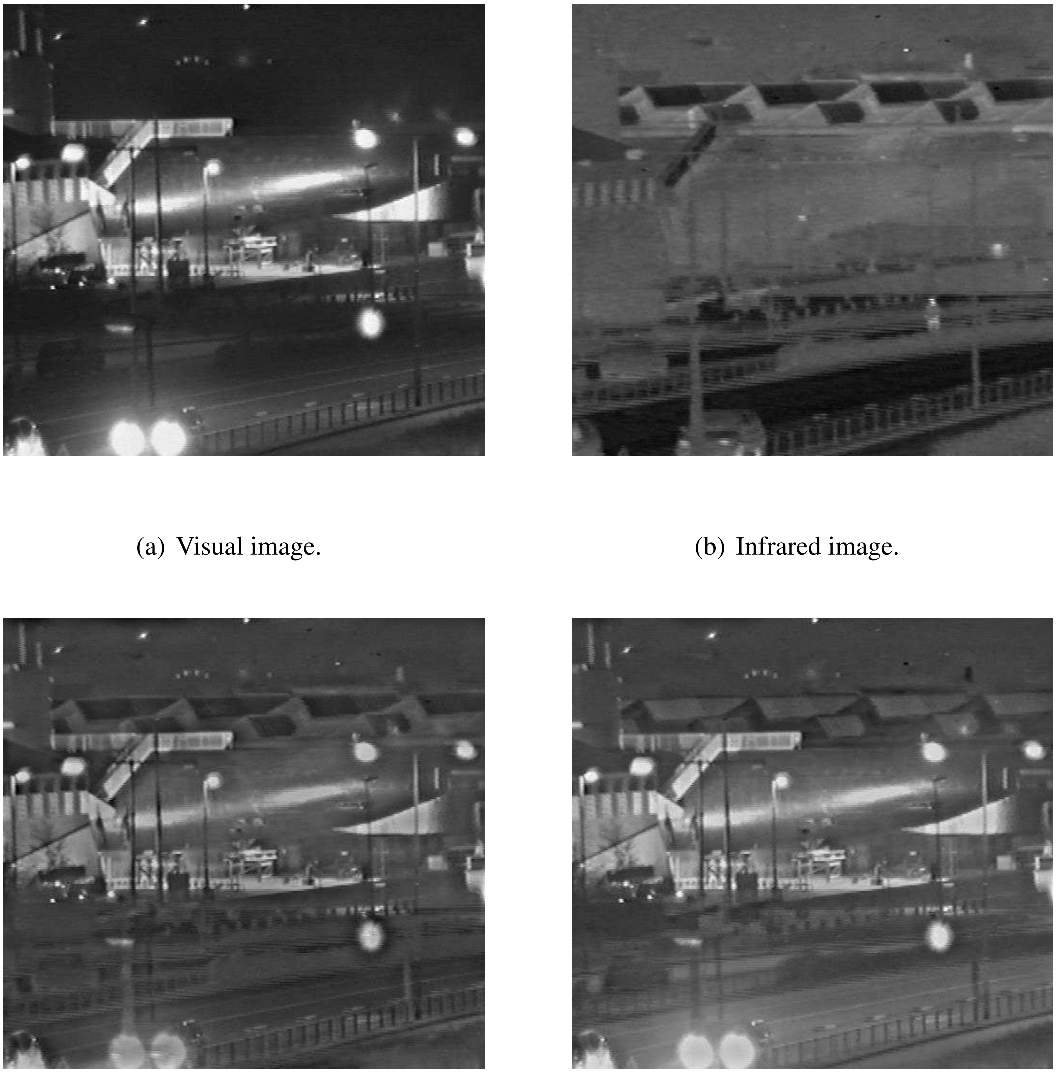
(b) Infrared image.

(c) MRA fusion result.

(d) Modified fusion result.

Figure 3.16: TNO Dune (frame 7404).

corresponding visual image, the pixels in the region of human body reach an almost uniform value of high intensity and the fusion process cannot reveal the insignificant difference in intensity. In this case, the image mosaic technique can be used to generate a composite image as described in [90]. The object (human body) needs to be identified from the IR image first. Then, the detected object can be embedded in the visual image. However,



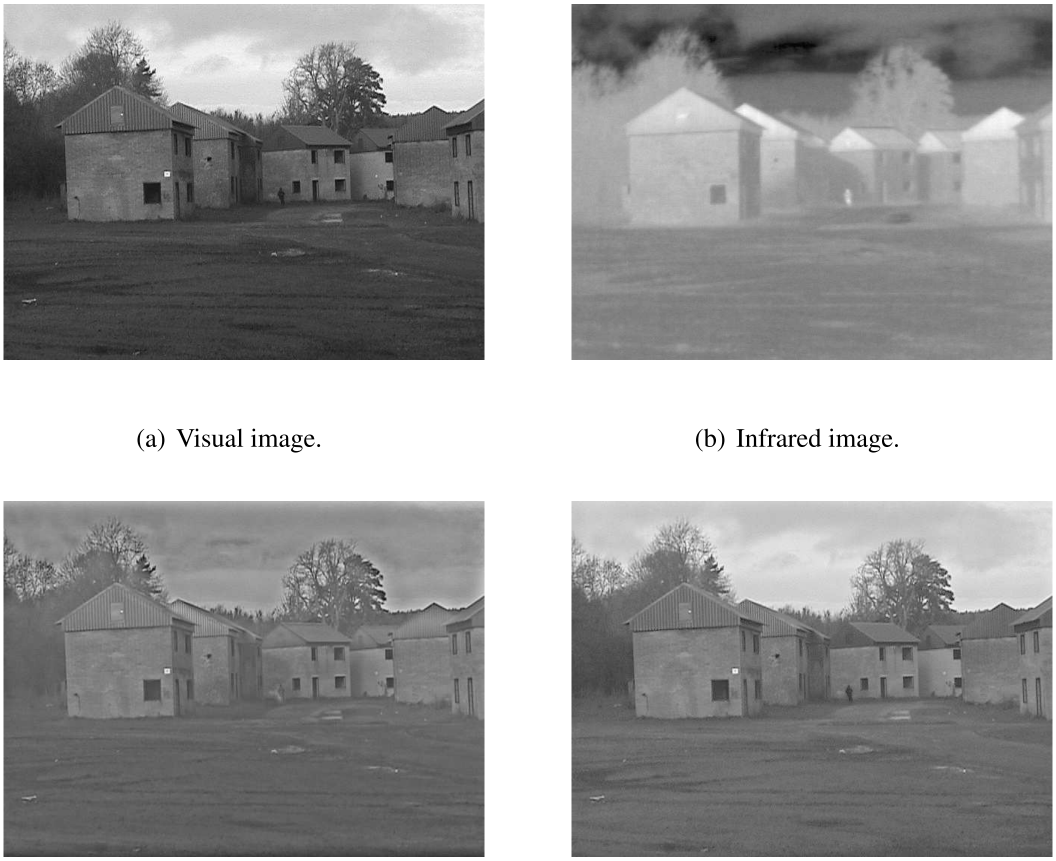
(a) Visual image.

(b) Infrared image.

(c) MRA fusion result.

(d) Modified fusion result.

Figure 3.17: TNO Kayak (frame e518a).



(a) Visual image.

(b) Infrared image.

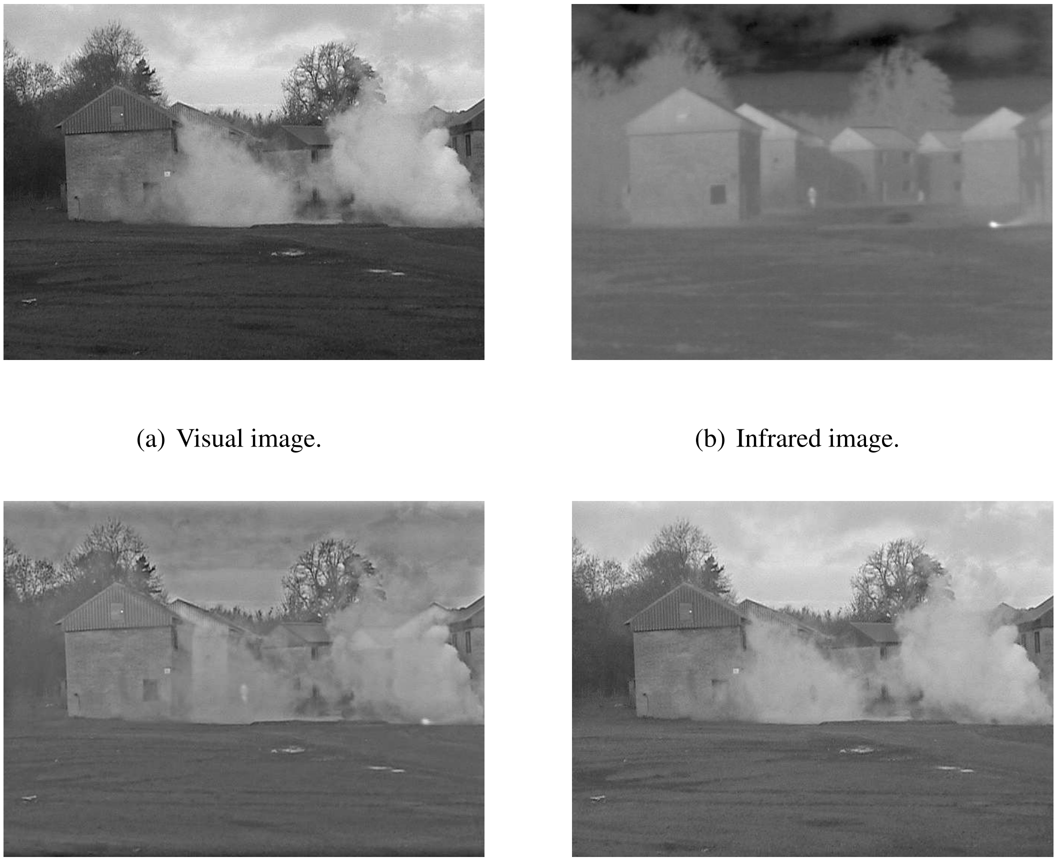
(c) MRA fusion result

(d) Modified fusion result.

Figure 3.18: Octec (frame 2).

this situation will not be a problem for the applications like surveillance or transportation, because the high-intensity pixels already aggregate a bright and distinct spot in the visual image, which appears to be a significant alert.

The IR image itself should be appropriate and good enough for the (human) object



(a) Visual image.

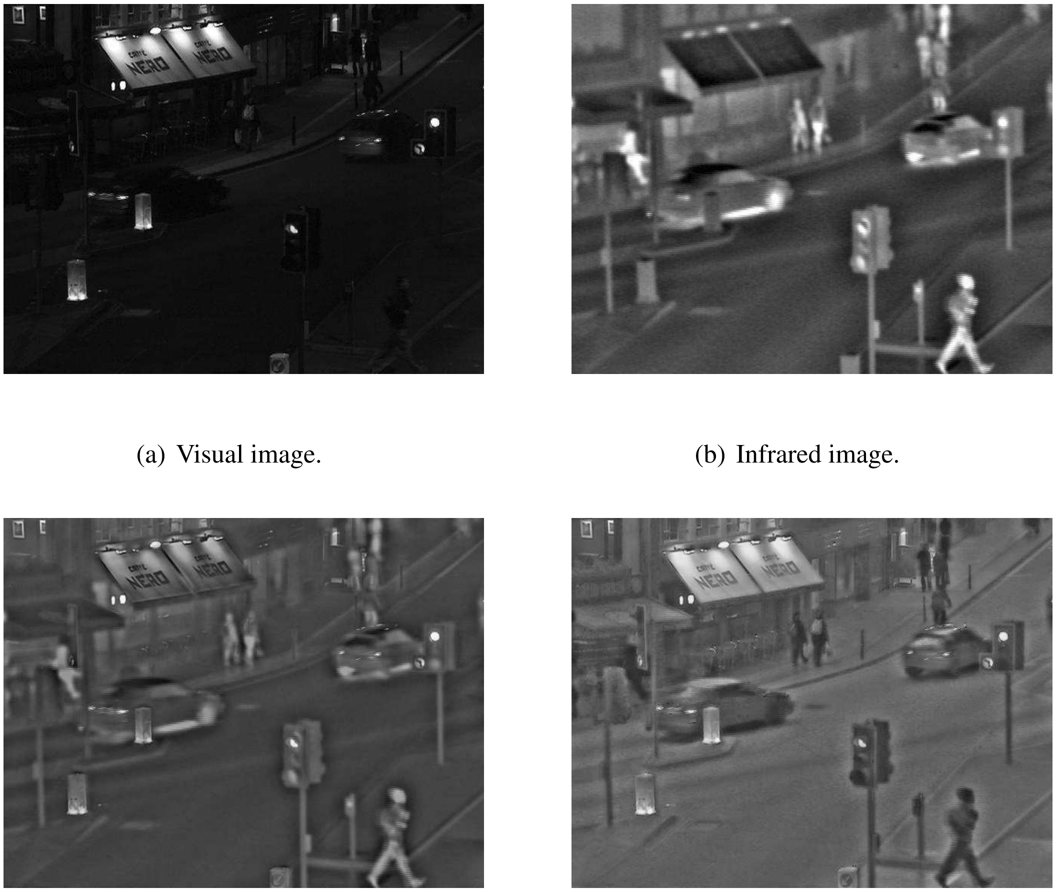
(b) Infrared image.

(c) MRA fusion result.

(d) Modified fusion result.

Figure 3.19: Octec (frame 21).

detection. The fusion with the visual image will not contribute to such purpose. However, when the task is to provide an observation of the scene, the fusion of the two image modalities as proposed in this chapter could be a good solution. Compared to the "direct" MRA-based pixel-level fusion, the results obtained with the proposed method are much



(a) Visual image.

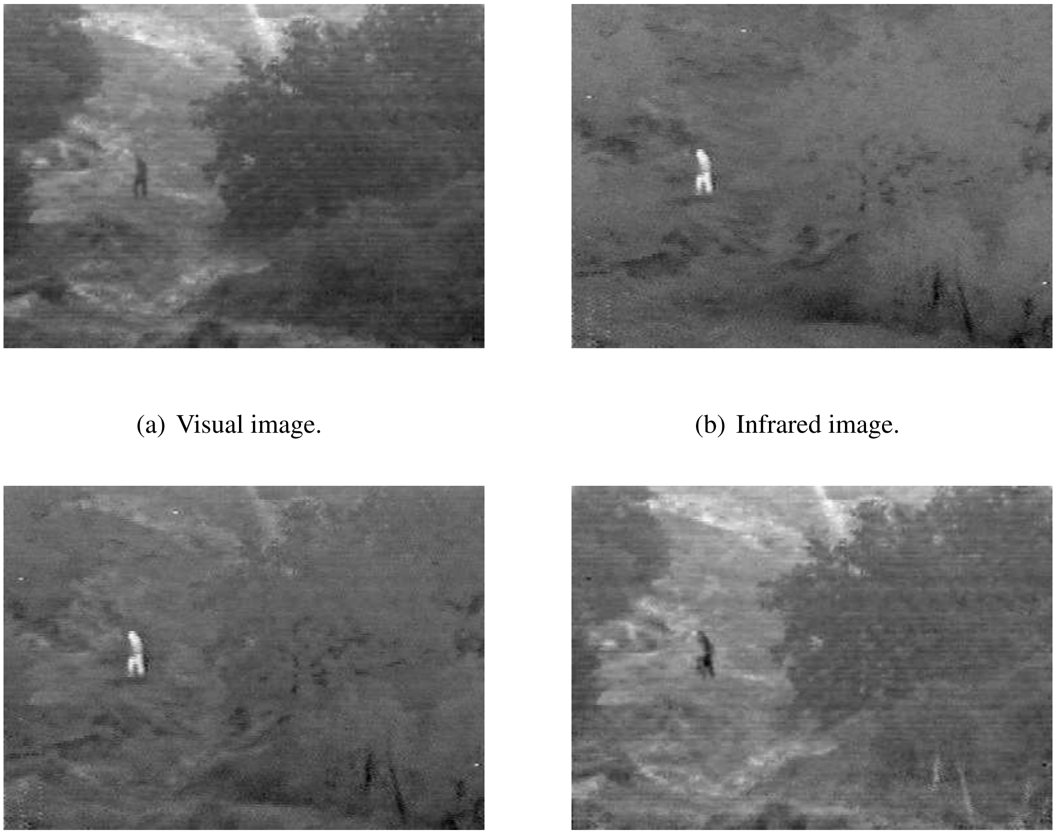
(b) Infrared image.

(c) MRA fusion result.

(d) Modified fusion result.

Figure 3.20: Bristol Queen’s road.

closer to the nature scene and therefore they are appropriate to human perception. In other words, the fusion results are still presented in the visual band. Future study may investigate if the fused image is useful for post-processing like segmentation in comparison with the results from IR image.



(a) Visual image.

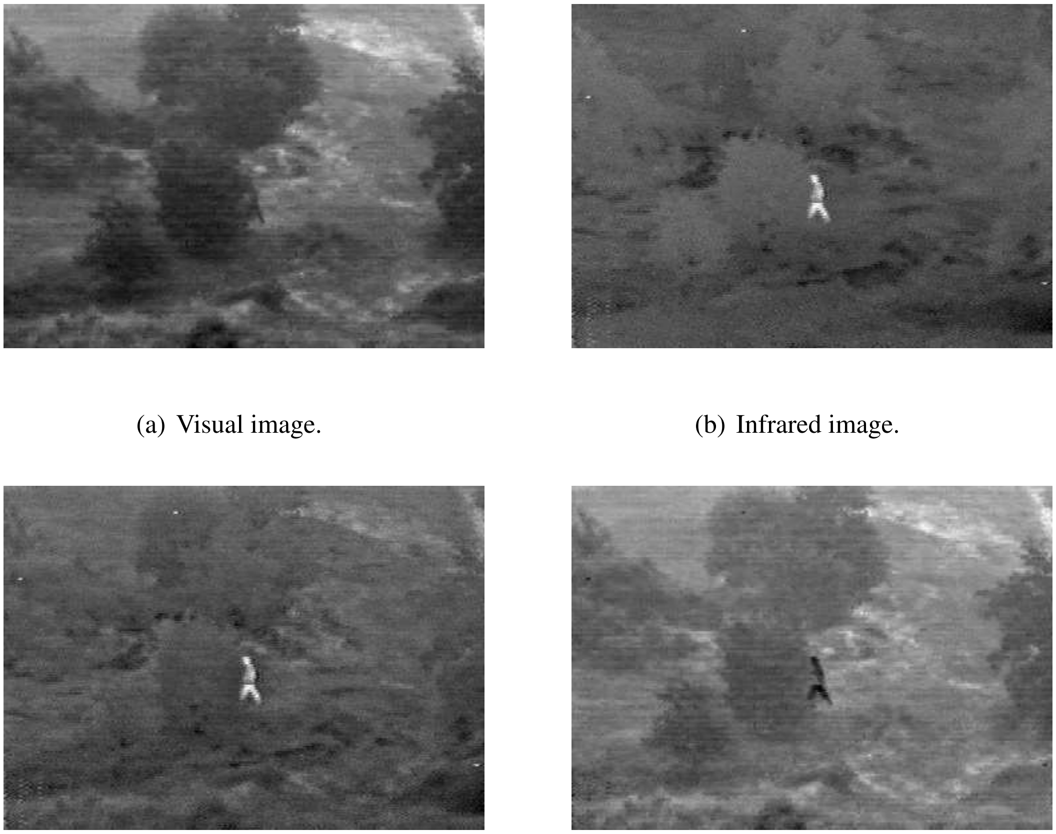
(b) Infrared image.

(c) MRA fusion result.

(d) Modified fusion result.

Figure 3.21: TNO trees (frame 4906).

Another important issue is the objective assessment of the efficiency of the fusion algorithms. This still remains a challenge for the research of image fusion, especially when no perfect reference image is available for comparison and this is the most case in a practical application. Current research focuses on the measurement of information transferred to the



(a) Visual image.

(b) Infrared image.

(c) MRA fusion result.

(d) Modified fusion result.

Figure 3.22: TNO trees (frame 4917).

fused image from the source images as described in Chapter 5 and 6. Experimental results on the evaluation will be presented in Chapter 6. Such measurement is valid for most cases. However, is a composite image, which contains the edges and boundaries from both IR and visual images, always the most optimal one for *human perception*, like the images in our experiments? Again, the answer is "no". The fusion operation should be able to convert the features from one spectrum band to another spectrum band rather than simply transfer those features. This could be the requirement for the applications where the fused images are presented for human perception instead of an automated processing or analysis procedure. Thus, the human factor issue needs to be considered.

In our method, the IR intensity value is simply used as the exponential function to enhance the visual image. More sophisticated functions may be considered in our future work, for example, the IR image can be segmented first and each segment can be applied with different functions upon the requirements.

3.7 Conclusion

In this chapter, a modified fusion process for night vision applications is presented. The method is straightforward and easy to implement. No empirical parameters need to be estimated and manually adjusted. Usually, a visual image is fused with the corresponding infrared image at the pixel level with MRA-based algorithms. However, the presentation of information is not optimal, because the features exhibiting in the range of visual band is more suitable for human perception. In the proposed method, the visual image obtained in an environment of poor visibility or inadequate illumination, is first enhanced by using the corresponding infrared image as the exponential factor. The enhanced result is then fused with the visual image to highlight the features in the visual spectrum band. In the fused result, the objects with higher radiation of heat are highlighted while the features from visual image are enhanced as well. This will help a driver to identify the obstacles or pedestrians on the road or improve the awareness of the environment in a surveillance application.

In the experiment, the fusion was implemented with the Simoncelli steerable pyramid. There are other options as described in section 3.2. The study on the evaluation of the fusion performance of different algorithms will be presented in Chapter 6.

Chapter 4

Concealed Weapon Detection and Visualization in a Synthesized Image

4.1 Introduction

THE image fusion technique provides a solution to combine information from multiple images and is able to generate a single image that gives a more accurate or complete description of the scene than any of the individual source images [5]. An example is the context enhancement presented in Chapter 3. However, there is no universal solution for all applications. Based on requirements of the concealed weapon detection application, there are different ways to implement the multi-modal image fusion process. The principle is illustrated in Figure 4.1. The first purpose of fusion is to facilitate the detection process. Like the circle in Figure 4.1, the fusion operation is to achieve an enhanced result to facilitate further analysis, recognition, or classification. Varshney et al. presented an automatic procedure to register and fuse infrared (IR) and millimeter wave (MMW) images in [9]. However, the study on how the further analysis can benefit from the fusion result is not available yet. The second purpose for CWD fusion is to locate human subjects with possible concealed weapons by fusing electro-optical (EO) and IR/MMW images [38]. Like the face and moon in Figure 4.1, the fused image contains both the personal information, i.e. facial pattern and the highlighted concealed weapon regions. This fusion is carried out at the pixel level as well. A human operator is presented with a composite image, with which the operator can respond accurately and promptly [38, 11, 60]. Another important issue has not been addressed yet is the "privacy rights". The multi-modal image device should not be used as a tool for voyeurism [91]. Therefore, the fusion algorithm must be tuned to reveal only the concealed weapon's information instead of personal privacy to the operators. The work presented in this chapter will focus on the second scenario, where a visual image is involved. In this chapter, the terminology "concealed weapon detection" is used to refer to the second topic aforementioned.

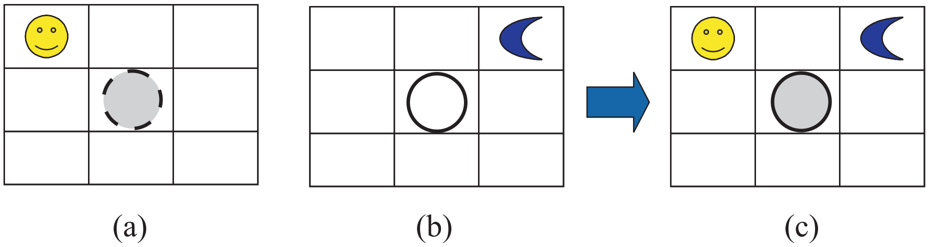


Figure 4.1: The illustration of image fusion techniques for concealed weapon detection applications. (a) and (b) are input images while (c) is the fusion result.

The philosophy of the approach presented in this chapter is different from previously published work, where a pixel-level fusion is carried out to the whole image. In this study, the weapon is first detected from an IR image by an unsupervised clustering algorithm, namely fuzzy k-means clustering. The feasibility of the clustering algorithm on IR or MMW image is investigated. The detected region is used as a mask signal for the multiresolution image mosaic (MRIM) process. The steerable pyramid is employed to decompose and reconstruct the two images. The reconstruction generates the final result. The rest of the chapter is organized as follows: A problem review is presented in section (4.2). A two-step scheme for synthesizing a composite image is described in section (4.3). Experimental results can be found in section (4.4). Discussion and conclusion are presented in section (4.5) and (4.6) respectively.

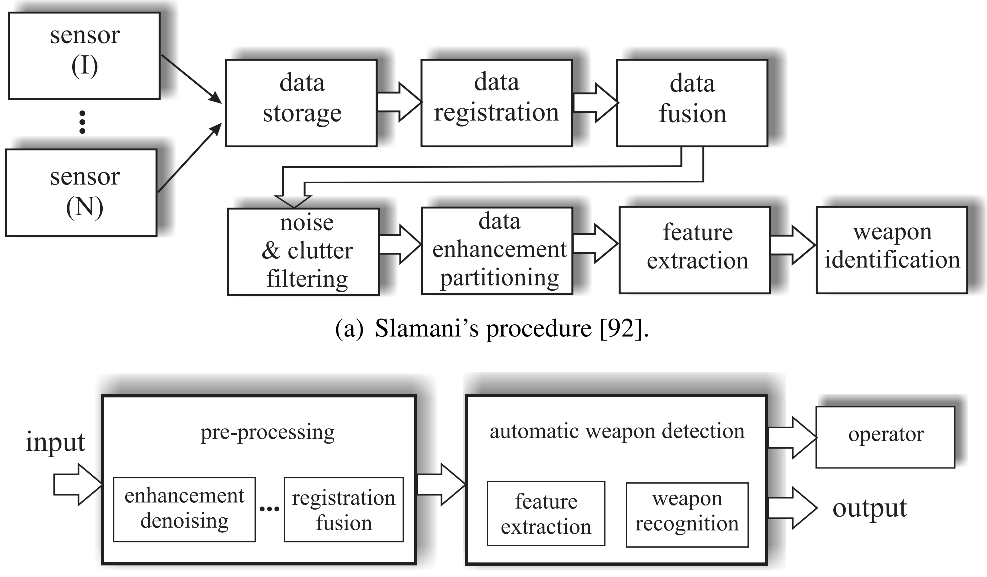
4.2 Problem Review

As IR-based CWD method is concerned, the basic principle is that the IR can image infrared radiation emitted by a human body, which is absorbed and re-emitted by clothing [38]. When there is a concealed weapon underneath the clothing, the radiation may vary. Thus, IR imaging can detect the concealed weapon and give the indication of its presence, when the clothing is tight, thin, and stationary. For normally loose clothing, the emitted IR radiation will be spread over a larger clothing area and causes the false alarm in the detection. Therefore, a longer wave length with a good penetration of clothing is preferred in such a scenario. In this chapter, we use the available IR images to test the algorithms. These algorithms can be applied to other nonvisual image data, for example, the MMW image.

To identify the procedure of processing CWD data, two schemes are presented as the flowcharts in Figure 4.2. The first one in Figure 4.2(a) was proposed by Slamani et al. in [92], where the filtering of noises was carried out after the fusion operation. The authors proposed another one (Figure 4.2(b)) in their recent publications [93, 38]. The second

[0] Concealed weapons do not have to use metallic material.

procedure is preferred in most cases, because the pre-processing needs to be applied before any further analysis is carried out. The pixel-level image fusion will retain salient features no matter whether these features are relevant or not. Such prominence will be presented in the final fused result and should be avoided.



(a) Slamani's procedure [92].

(b) Vashney's procedure [93].

Figure 4.2: The signal processing procedures for CWD.

Slamani et al. also proposed a mapping procedure consisting of three stages in [94]. The first stage is threshold computation, which segments the original image into a number of binary scenes. A low-pass filter and a high-pass filter are used to group pixels and detect edges for each scene in the second stage. At the third stage, a composite is obtained by summing all the processed sub-images together. This procedure actually accomplished a clustering of pixels with common features and will directly affect the systematic performance.

The fusion of IR and MMW images has been studied by Salmani [92] and Varshney [9] respectively. In [7] and [6], Uner and Slamani fused multiple IR images with a discrete wavelet transform. In [11], Xue and Blum did an extensive study on fusion of visual and IR images with different MRIF algorithms. The fused results were evaluated by a number of quantitative metrics. However, the visual quality of the fused image was degraded in most of the experimental results. The problem is that the MRA algorithms try to keep the salient features of images no matter whether the substance is really useful or not. The disadvantage of the MRIF approach is that when the two source images are of great difference, the selecting or even the averaging of the low pass components will cause the “block” effect in the fused result. In other words, the reconstruction is not stable. Lately Xue presented a new color-based fusion algorithm, in which IR image is fused with color channels [60]. Yang et al. employed the expectation-maximization algorithm to estimate the optimal scene in [58].

As described in [38], the further processing is towards an automatic weapon detection. Commonly used object extraction approaches are based on thresholding or segmentation techniques. In Slamani’s mapping procedure A’SCAPE [6], homogeneous regions are separated by applying a series of threshold values followed by a low- and high-pass filtering operation. The basic idea is to group pixels in homogeneous regions. In [9], the authors suggested using Otsu’s thresholding method [95] to the fused result of IR and MMW images. However, there is no study on assessing the performance of these approaches so far. Current available fusion techniques for a CWD application are summarized in Table 4.1 and the details will not be repeated herein.

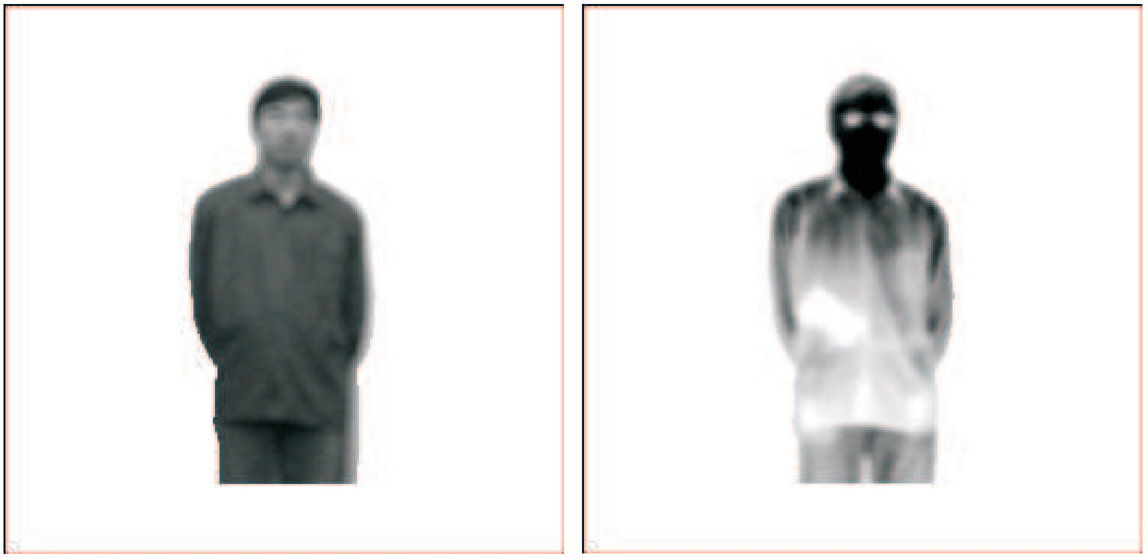
Table 4.1: The summary of the image fusion techniques for CWD.

Image modality	Method	Achievement	References
fusion of two IR images	spline wavelet transform and Burt's fusion rule [5]	obtain more complete and detailed information	Üner [7], Slamani [6]
fusion of IR and MMW images		facilitate the shape extraction process	Slamani [92], Varshney [9]
fusion of IR and visual images	comparison of 15 MRA fusion algorithms	retain the fidelity of facial pattern and highlight the concealed weapons	Xue [11]
	color-channel fusion		Xue [60]
	expectation maximization (EM) algorithm		Yang [58]
	EM and hidden Markov model		Yang [96]
	region-based EM algorithm		Yang [97]
	image mosaic		Liu, Blum [90, 98]

An example for CWD is shown in Figure 4.3. Picture on the left is the visual image while the corresponding IR image is on the right. For visualization purpose, the inverse image is used, i.e. the higher intensity value corresponds to the lower temperature point. The concealed weapon can be visually detected from the IR image. Current study on multiresolution image fusion for CWD is to generate a composite image for the operator or an automated analysis procedure as shown in Figure 4.4(a). We suggest a new image processing framework in Figure 4.4(b). Each pixel from the IR and/or MMW images is classified with a confident value as belonging to either a weapon or a non-weapon region.

[0] The input images are assumed to be fully registered.

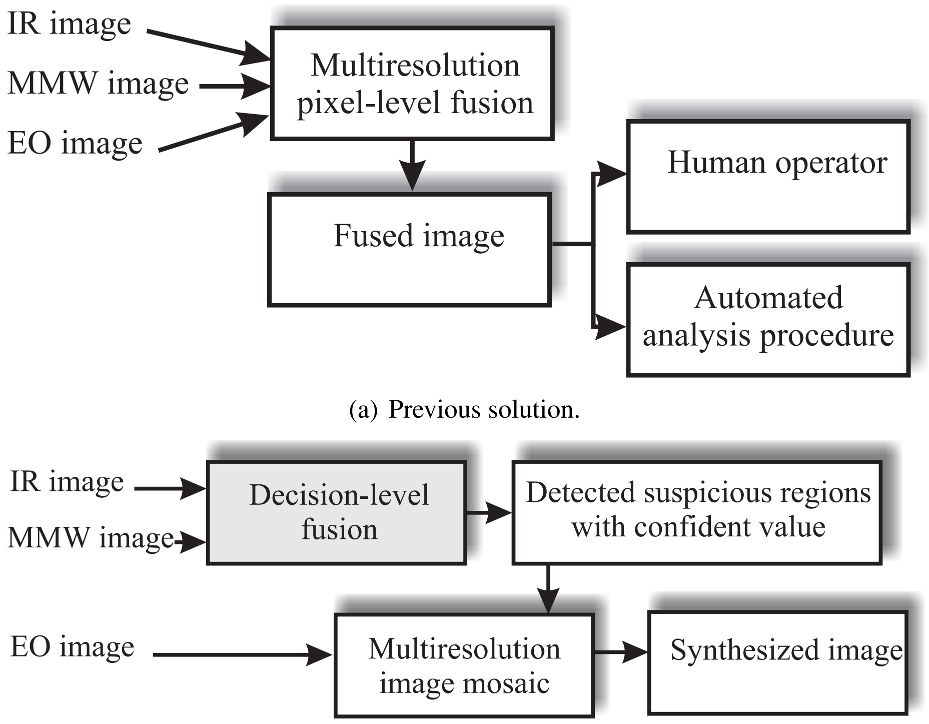
This can be implemented at a higher level (decision level instead of pixel level). The detected region is further segmented by a predefined confidence threshold and embedded into the corresponding visual image by using a multiresolution image mosaic (MRIM) technique, which can achieve a seamless boundary between host image and embedded regions. In this work, only the selected (weapon region) parts are synthesized with the visual image, because other parts does not contribute to the weapon detection at all.



(a) Visual image.

(b) Infrared image.

Figure 4.3: An example of image pair for CWD.


IR image
MMW image
EO image
Multiresolution pixel-level fusion
Fused image
Human operator
Automated analysis procedure
IR image
MMW image
Decision-level fusion
Detected suspicious regions with confident value
EO image
Multiresolution image mosaic
Synthesized image

(a) Previous solution.

(b) Proposed method.

Figure 4.4: The image processing architectures for CWD applications..

4.3 A Two-Step Scheme for Synthesizing a Composite Image

The objective of synthesizing a visual and non-visual image is to retain the information of both the personal identification and the concealed weapons. It is obvious that the IR image contributes little to the facial identification in the case of being fused with a visual image. Therefore, a simple combination may degrade the quality of the fusion result for facial identification. The detection of concealed weapon depends on the operation of the infrared sensor, because the pixel value of the IR image reflects the variations in temperature. If the infrared sensor cannot find out the concealed weapon, the fusion with a visual image will not generate a useful result. The temperature variance of different objects, i.e. weapon, clothing, and body, can be identified by using an unsupervised clustering approach. A two-step scheme consisting of a detecting and an embedding operation is proposed next.

4.3.1 Concealed Weapon Detection

Fuzzy k-means Clustering

Fuzzy k-means clustering assigns a membership grade to a data point belonging to certain cluster [99]. It is an unsupervised approach for data clustering through seeking a minimum of heuristic global cost function [99]:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} \left[\hat{P}\left(\omega_i \middle| x_j, \hat{\theta} \right) \right]^b (x_j - \mu_i)^2 \tag{4.1}$$

where the probability $\hat{P}\left(\omega_i \middle| x_j, \hat{\theta} \right)$ stands for the fuzzy membership of pixel x_j $(j = 1 \cdots n)$ in a cluster ω_i $(i = 1 \cdots c)$, and there are in total c clusters in the data set. The mean value for each cluster ω_i is μ_i. The number b is a free parameter chosen to adjust the blending of different clusters, while $\hat{\theta}$ is the parameter vector for the membership functions. The probabilities of cluster membership for each pixel are normalized as:

$$\sum_{i=1}^{c} \hat{P}\left(\omega_i \left| x_j \right. \right) = 1, \ j = 1, \cdots\cdots, n \tag{4.2}$$

The minimization of the cost function in equation (4.1) leads to the solutions [99]:

$$\mu_j = \frac{\sum_{j=1}^{n} \left[\hat{P}\left(\omega_i \left| x_j \right.\right)\right]^b x_j}{\sum_{j=1}^{n} \left[\hat{P}\left(\omega_i \left| x_j \right.\right)\right]^b} \tag{4.3}$$

and

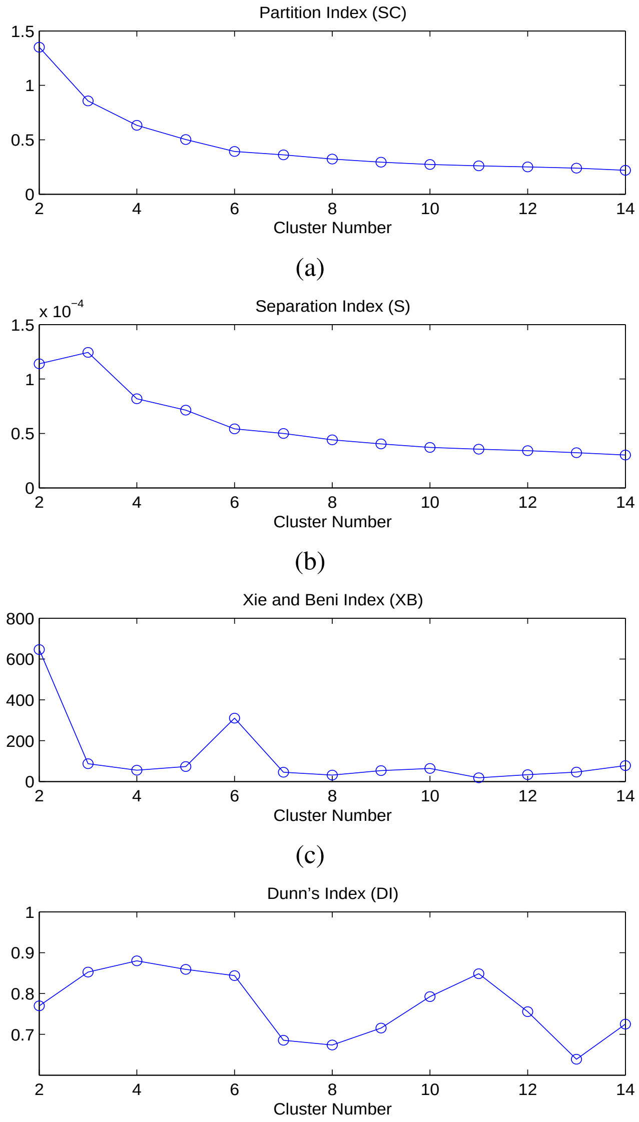
$$\hat{P}\left(\omega_i \left| x_j \right.\right) = \frac{\left(1/d_{ij}\right)^{1/(b-1)}}{\sum_{r=1}^{c} \left(1/d_{rj}\right)^{1/(b-1)}} \; and \; d_{ij} = \left(x_j - \mu_i\right)^2 \tag{4.4}$$

The cluster means and point probabilities are estimated iteratively until there is only small change in μ_j and $\hat{P}\left(\omega_i \left| x_j \right.\right)$.

By applying the fuzzy k-means clustering algorithm to the IR images, a set of clustered images are obtained. The idea is similar to Slamani's SMP [92] in grouping pixels in homogeneous regions. It is observed that the cluster corresponding to the highest center value is the collection of the points in the concealed weapon region. By applying a proper threshold value, a binary mask image is obtained and used for the mosaic operation.

The fuzzy k-means clustering algorithm needs the number of clusters as an input parameter, which can be determined empirically. Calculating validity measure indexes can help to estimate the goodness of the fuzzy clustering algorithm and find the optimal number of clusters [100]. Herein, four validity indexes are employed, i.e. partition index(SC), separation index(S), Xie and Beni's index(XB), and Dunn's index(DI) [100]. Detailed description and implementation of these metrics are available in [101, 102, 103, 100].

A small cluster number is better for computational efficiency. The clustering accuracy should be considered when the cluster number is determined. In Figure 4.5, SC and S index hardly decrease at point 8 while XB and DI reach their local minimum at the same point. Therefore, in our experiments, we select eight as the initial number of clusters for the IR images.



(a)

(b)

(c)

(d)

Figure 4.5: The clustering indexes (a) partition index, (b) separation index, (c) Xie & Beni index, and (d) Dunn's index, with different cluster numbers.

Region of Interest (ROI) Enhancement

The aforementioned approach provides another advantage that particular processing can be applied to the region of interest partitioned by the mask image. On one hand, the synthesized image is evaluated by the operator; on the other hand, in the further processing, different algorithms can be applied to the different ROI regions respectively. For example, if we again apply the fuzzy k-means clustering algorithm to the ROI of an IR image, the shape of the weapon can be detected through finding out the cluster with the highest center value. With this information, the weapon in the IR image can be enhanced. If only the shape is enhanced, we can simply multiply the IR image with a gain map in which the value in the weapon region is larger than 1. Another enhancement scheme is to use the corresponding membership map from the IR image:

$$I_{IR}\left(x,y\right)=I_{IR}\left(x,y\right)\left(1+\alpha\cdot F_{ROI}\left(x,y\right)\right) \tag{4.5}$$

$F_{ROI}\left(x,y\right)$is the corresponding ROI fuzzy membership map. The pixel with the higher membership value is emphasized more by the parameter α. The next step is to follow previously descibed procedure to mosaic the visual image and the enhanced IR image.

4.3.2 Embedding in a Visual Image

The idea of multiresolution image mosaic is to combine two or more images into a composite one with an invisible seam [5, 104]. The general procedure is shown in Fig. 4.6. Like the multiresolution image fusion process, the input images are decomposed by a certain multiresolution algorithm Ψ. Meanwhile, the Gaussian pyramid of the binary mask image is constructed $GI_N\left(x,y\right),\cdots GI_2\left(x,y\right)$,$GI_1\left(x,y\right)$, where N is the decomposition level. The new image components can be formed by the weighted sum with the Gaussian image components. There are several ways to achieve this.

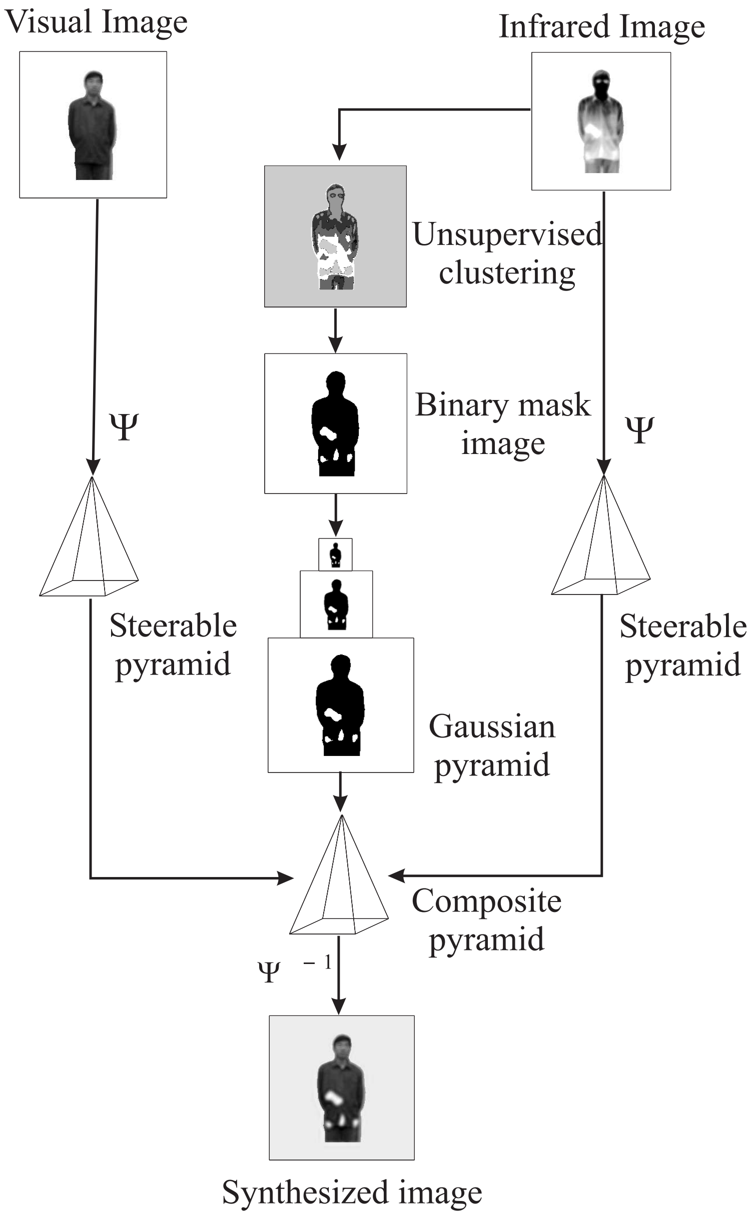


Figure 4.6: The procedure for multiresolution image mosaic.

The first implementation is achieved by the weighted summation of every image components. Recall the representation of a decomposition with the steerable pyramid:

$$I\left(x,y\right) \to \left(\; LI\left(x,y\right), \;\; BI_i^j\left(x,y\right)\Big|_{i=1\ldots N}^{j=1\ldots K}, \;\; HI\left(x,y\right) \; \right) \tag{4.6}$$

The low-pass, band-pass, and high-pass component are $LI(x,y)$, $BI_i^j(x,y)$, and $HI(x,y)$ respectively. The formulae are given below:

$$\begin{aligned} HI\left(x,y\right) &= GI_1\left(x,y\right) \cdot HI_{IR}\left(x,y\right) + \left(1 - GI_1\left(x,y\right)\right) \cdot HI_V\left(x,y\right) && (4.7) \\ BI_i^j\left(x,y\right) &= GI_i\left(x,y\right) \cdot BI_{IRi}^j\left(x,y\right) + \left(1 - GI_i\left(x,y\right)\right) \cdot BI_{Vi}^j\left(x,y\right) && (4.8) \\ LI\left(x,y\right) &= GI_N\left(x,y\right) \cdot LI_{IR}\left(x,y\right) + \left(1 - GI_N\left(x,y\right)\right) \cdot LI_V\left(x,y\right) && (4.9) \end{aligned}$$

The new image components will be used to reconstruct the composite image. The second implementation uses the edge information of the mask image map. The original edge map can be easily obtained by the Canny edge detector. Instead of generating a Gaussian image pyramid, through the down-sampling operation, we can get a set of edge images $E_N\left(x,y\right), \cdots E_2\left(x,y\right), E_1\left(x,y\right)$ and mask images $M_N\left(x,y\right), \cdots M_2\left(x,y\right), M_1\left(x,y\right)$. Now, the combination formulas become:

$$\left\{ \begin{array}{cc} I_{IR}\left(x,y\right) & M\left(x,y\right) = 1, E\left(x,y\right) = 0 \\ \left(I_{IR}\left(x,y\right) + I_V\left(x,y\right)\right)/2 & M\left(x,y\right) = 1, E\left(x,y\right) = 1 \\ I_V\left(x,y\right) & M\left(x,y\right) = 0, E\left(x,y\right) = 1 \end{array} \right. \tag{4.10}$$

$$LI\left(x,y\right) = M_N\left(x,y\right) \cdot LI_{IR}\left(x,y\right) + \left(1 - M_N\left(x,y\right)\right) \cdot LI_V\left(x,y\right) \tag{4.11}$$

The operation will copy the corresponding regions from the visual and IR images to the new image component, i.e. “cut and paste”. At the edge between the two regions, an average operation is applied. In above equations, $I_{IR}\left(x,y\right)$, $I_V\left(x,y\right)$, and $I\left(x,y\right)$ stand for the high- and band-pass image components of IR, visual, and new images respectively. For the low-pass component, we do not use the edge to smooth the transition zone. The discussion can be found in section 4.5. The third implementation differs from the first in the combination of low pass components. For the high-pass and band-pass components, equation (4.7) and (4.8) are applied. The low pass component from the visual image is retained as the new low pass component for reconstruction; or a weighted summation is implemented in the marked weapon region by the mask image map. Such operations can also be applied for texture mapping [104].

4.3.3 Result Assessment

An ideal solution for evaluating the fused image is to compare it with a reference image, which is assumed to be perfect. However, such reference image is not available in advance for the CWD application. The success of the application largely depends on whether the suspicious regions can be detected or not. Therefore, the classification metrics, accuracy and reliability, are employed herein. An illustration to interpret this concept is given in Fig. 4.7. Suppose A is the ground truth (true weapon region), B is the detected result (detected weapon region) and C is the overlap between A and B. The accuracy is defined as the ratio between the positively true and all pixels that are used as the ground truth of this class, i.e. $\frac{C}{A} \times 100\%$, while the reliability is expressed as $\frac{C}{B} \times 100\%$, i.e. the ratio between the positively true and all pixels classified as this class. A large accuracy value together with a higher reliability indicates a good classification result.

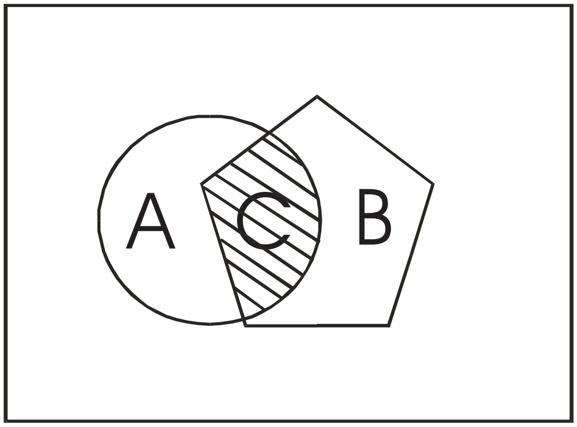
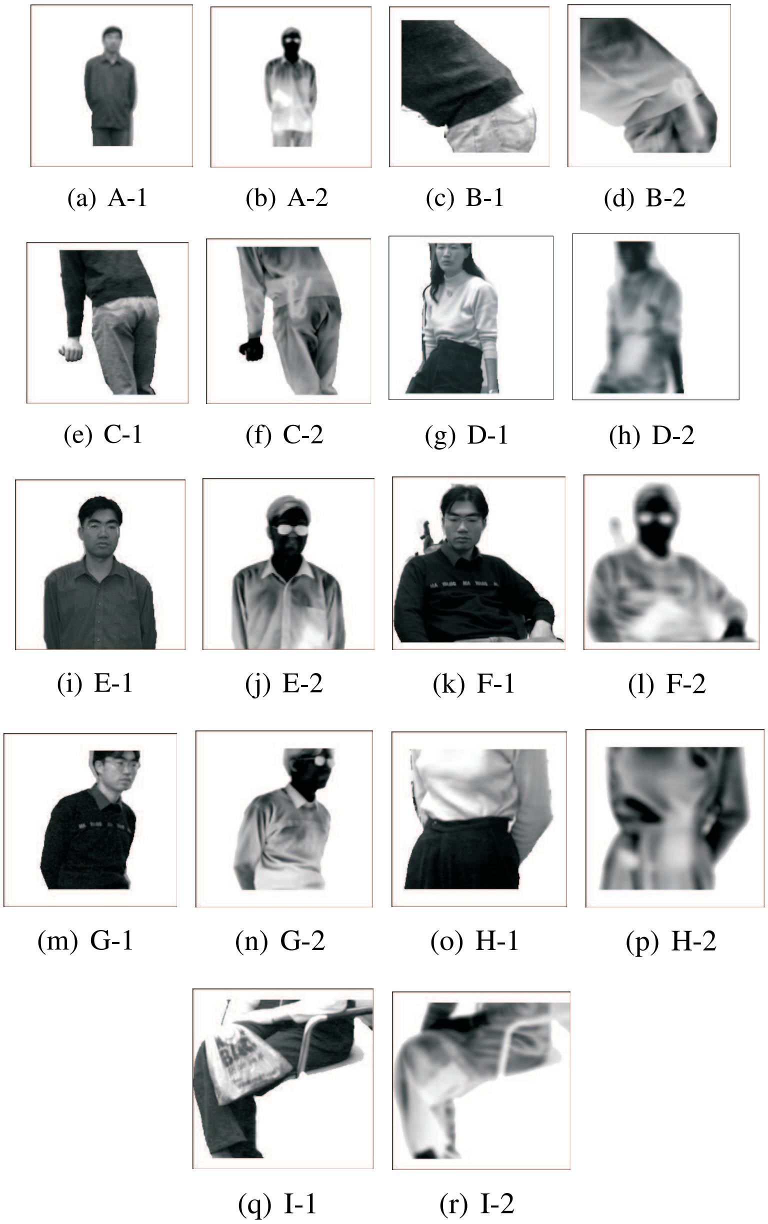


Figure 4.7: Illustration for accuracy and reliability assessment.

4.4 Experimental Results

The multi-sensor image data was collected at the Signal Processing and Communication Laboratory (SPCL) of Lehigh University. There are nine pairs of visual and IR images shown in Figure 4.8. In the following experiments, we assume: 1) the visual image and IR image are fully registered; 2) both the visual and IR image are background subtracted; and 3) there is a concealed weapon in each scene.

In the first part of the experiment, the first pair of images in Figure 4.8 was integrated by image fusion algorithms. Figure 4.9 presents the results obtained by applying Laplacian pyramid, Daubechies wavelet and Simoncelli steerable pyramid based fusion algorithms respectively. The coefficient combination rule is: averaging the low pass image components and applying the maximum selection rule to the high pass components. More sophisticated rules and algorithms were implemented in [5, 70, 63]. The steerable pyramid based algorithm was presented in [69] and applied to the image pair in Figure 4.8(a) and 4.8(b).



(a) A-1 (b) A-2 (c) B-1 (d) B-2

(e) C-1 (f) C-2 (g) D-1 (h) D-2

(i) E-1 (j) E-2 (k) F-1 (l) F-2

(m) G-1 (n) G-2 (o) H-1 (p) H-2

(q) I-1 (r) I-2

Figure 4.8: Multi-sensor images used for testing in the experiment: totally eight groups are involved (A-I).

Figure 4.9(c) and 4.9(d) give the results. The facial pattern is obscure in the pixel-level fusion results, although the weapon region can be observed to some extent.

In the second part of the experiments, the multiresolution image mosaic was implemented. As described in section 4.3.2, there are three approaches that come with the multiresolution mosaic scheme. To apply the mosaic algorithm, the mask signal needs to be extracted. In Figure 4.10(a), the segmented result by applying fuzzy k-means clustering algorithm is shown. By selecting the cluster with the highest center value and applying a proper threshold value, the binary image map was obtained and given in Figure 4.10(b). In the experiment, the points in this cluster with a fuzzy membership value larger than 0.1 were collected and averaged. The averaged value was selected as the threshold. With the binary mask image, the visual and IR images were synthesized by the proposed algorithms. The decomposition level of the multiresolution representation does affect the results. We gave the results with two, three, and four level decomposition in Figure 4.11.

Table 4.2: Comparison of the Fuzzy k-means clustering results with different initial cluster number.

Cluster number	8	10	13	16	19	22	25	30	40
False positive	0.5146	0.4314	0.3718	0.3444	0.3014	0.3014	0.2819	0.2552	0.2552
True positive	1	1	0.9721	0.9604	0.9249	0.9249	0.9106	0.901	0.901

To see how the number of clusters affects the detection of weapon region in terms of accuracy and reliability measurements, we used a set of numbers in Table 4.2 to cluster IR image of Figure 4.8(b) and compared the detected results with a manually generated reference image. Figure 4.12 shows the curve. A larger cluster number can achieve a higher reliability at the cost of losing accuracy; meanwhile, a larger number will introduce

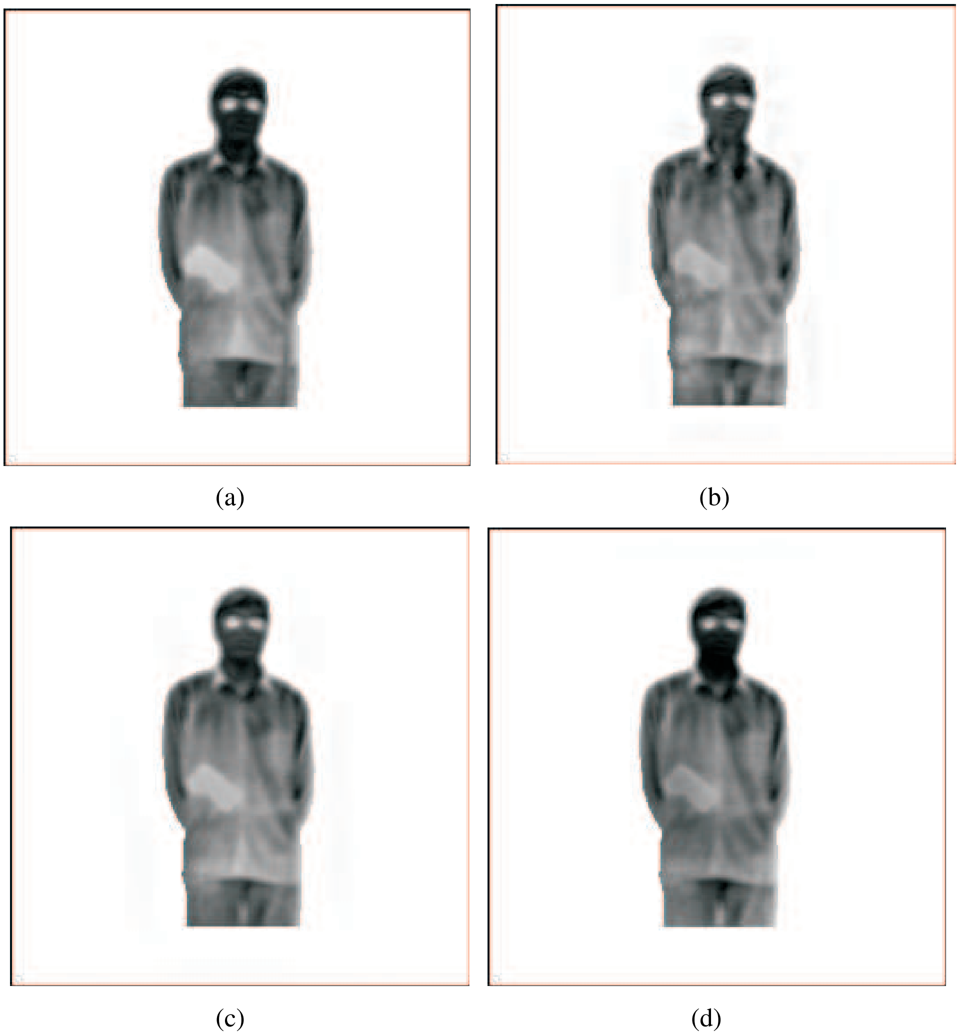
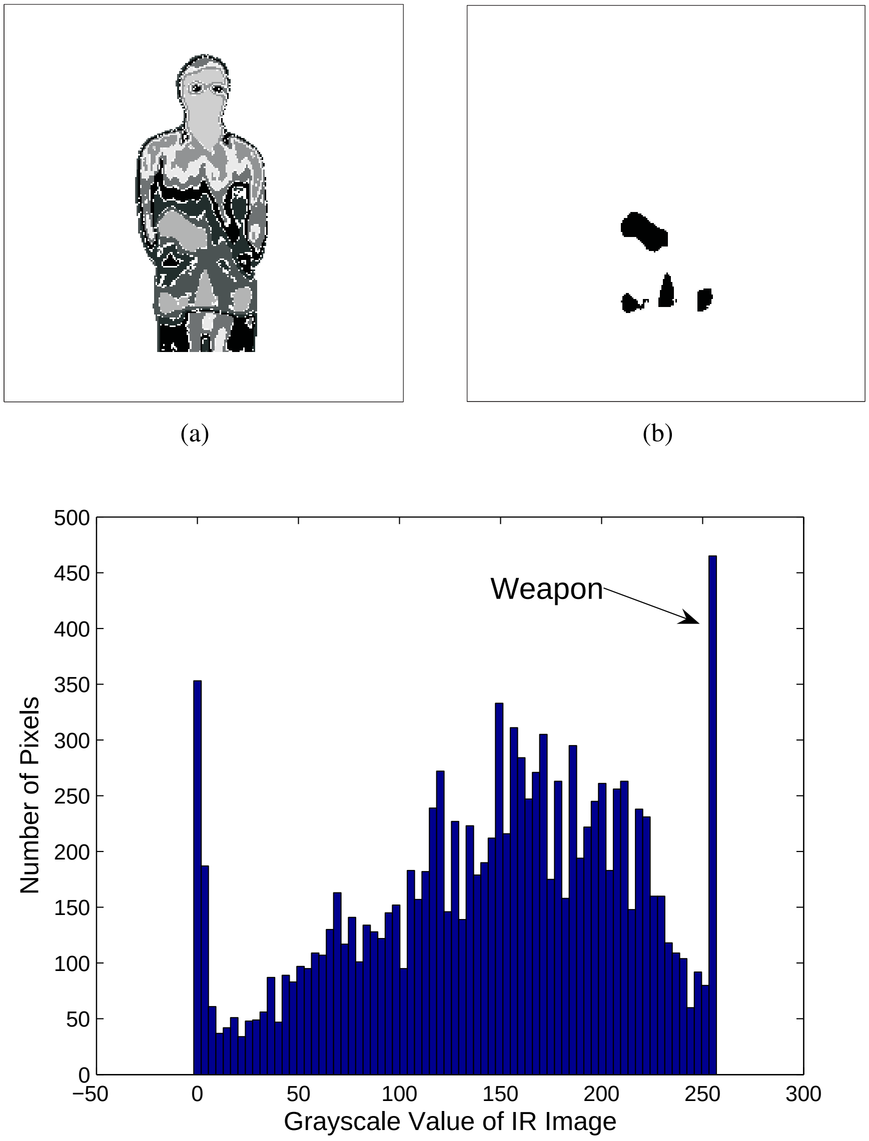


Figure 4.9: Image fusion results achieved by (a) Laplacian pyramid; (b) Daubechies wavelet four; (c) Simoncelli steerable pyramid (averaging for low-pass component and maximum selection for band- and high-pass components); and (d) Simoncelli steerable pyramid with sub-band images integrated by Laplacian pyramid).



(a)

(b)

(c)

Figure 4.10: (a) Clustered image by fuzzy k-means clustering algorithm; (b) binary mask image obtained from the clustered result; and (c) histogram of IR image.

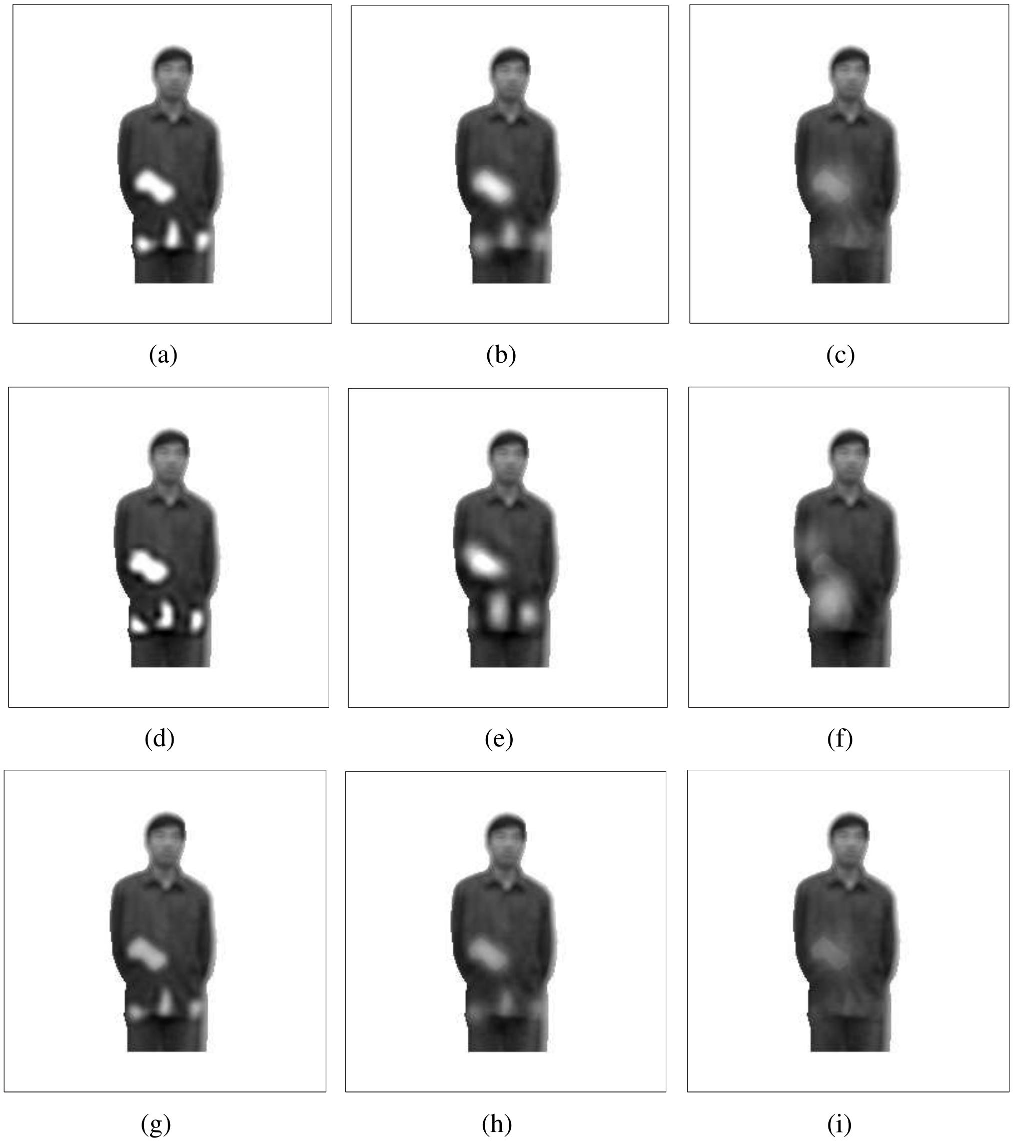


Figure 4.11: Mosaic results achieved by applying the multiresolution approach one at different decomposition level (a) 2, (b) 3, and (c) 4; approach two at decomposition level (d) 2, (e) 3, and (f) 4; approach three at decomposition level (g) 2, (h) 3, and (i) 4.

computational loads. For the CWD application, a higher accuracy has the priority over the reliability in most cases.

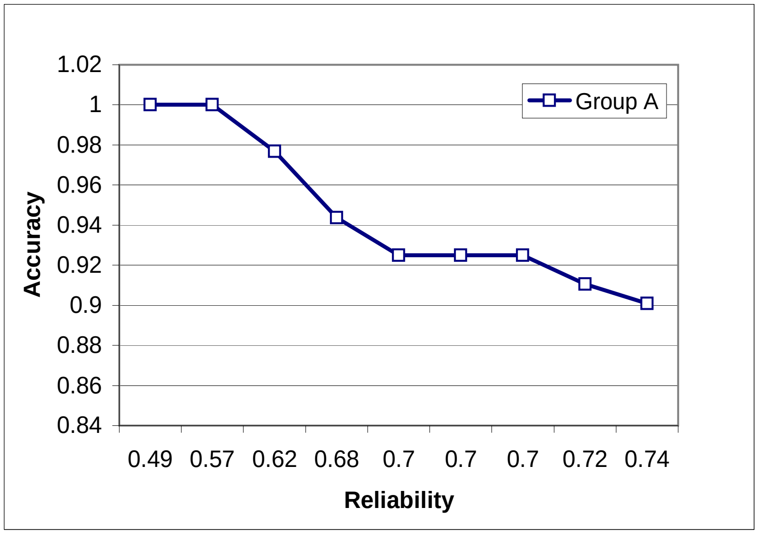


Figure 4.12: The effect of cluster number for IR image of Group A in Figure 4.8(b)

.

In addition, we compared the fuzzy k-means clustering method with the expectation-maximum (EM) clustering and k-means clustering methods. The three clustering algorithms were applied to the nine groups of multi-sensor images with the same cluster number 8. The results of accuracy and reliability assessments are listed in Table 4.3 and illustrated in Figure 4.13.

In terms of classification rate, the fuzzy clustering does not show obvious advantages over the other approaches. Nevertheless, the outputs of fuzzy clustering can be used to enhance the region of interest (ROI) in the IR image. The concealed weapons in Figure 4.8(d) and 4.8(f) have explicit shapes. The enhancement may facilitate the further processing.

Table 4.3: Comparison of clustering schemes.

	Fuzzy k-means Clustering		EM Clustering		k-means Clustering	
	Accuracy	Reliability	Accuracy	Reliability	Accuracy	Reliability
Group A (1)	1	0.4854	1	0.2344	1	0.3917
Group B (2)	1	0.4569	1	0.3546	1	0.3828
Group C (3)	0.9529	0.4868	0.9532	0.4455	0.9540	0.3329
Group D (4)	0.4336	0.4117	0.4946	0.4077	0.5373	0.4090
Group E (5)	0.8618	0.6695	0.8618	0.5431	0.8618	0.6217
Group F (6)	0.9254	0.5066	0.9254	0.3414	0.9254	0.3290
Group G (7)	0.9776	0.8104	0.9776	0.6150	0.9776	0.6539
Group H (8)	0.4767	0.5556	0.5100	0.5055	0.8211	0.4980
Group I (9)	0.2248	0.2412	0.3222	0.2898	0.3895	0.3117

First, we used the binary mask image to extract the ROI of the IR image. Then, the ROI was segmented again by the clustering algorithm. The region of the concealed weapon was further refined. By using the fuzzy membership map of the ROI, the IR image can be enhanced according to equation (4.5). The visual image was then synthesized with the enhanced version of the IR image. Figure 4.14 and 4.15 show the results.

From the above experiments, we can see that the third multiresolution mosaic approach with a decomposition level two achieved a better result in terms of human perception, i.e. a subjective assessment. Eventually, we applied this approach to the other images and give the results in Fig. 4.16.

4.5 Discussion

The advantages of pixel-level fusion of IR and MMW images are not explicitly identified; therefore, a decision-level fusion for classification is suggested. In this study, we did

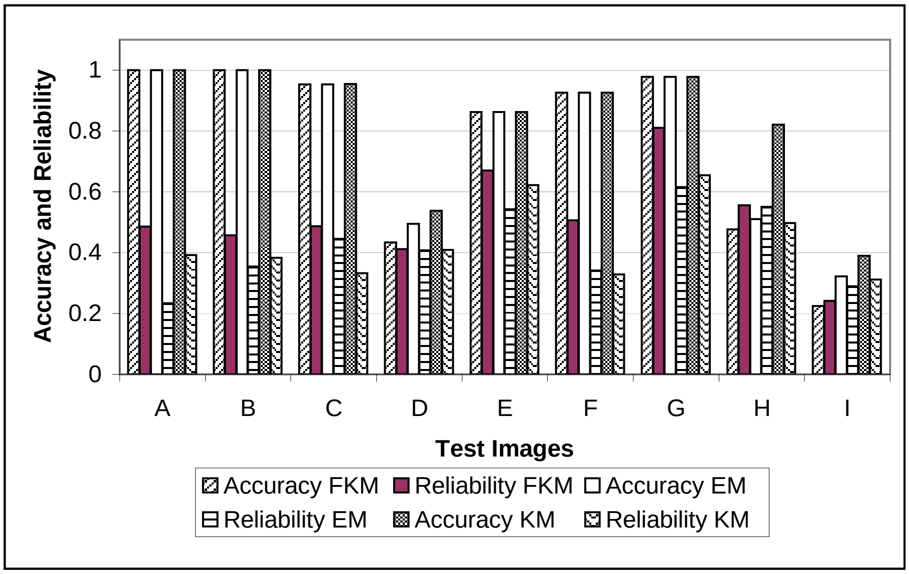
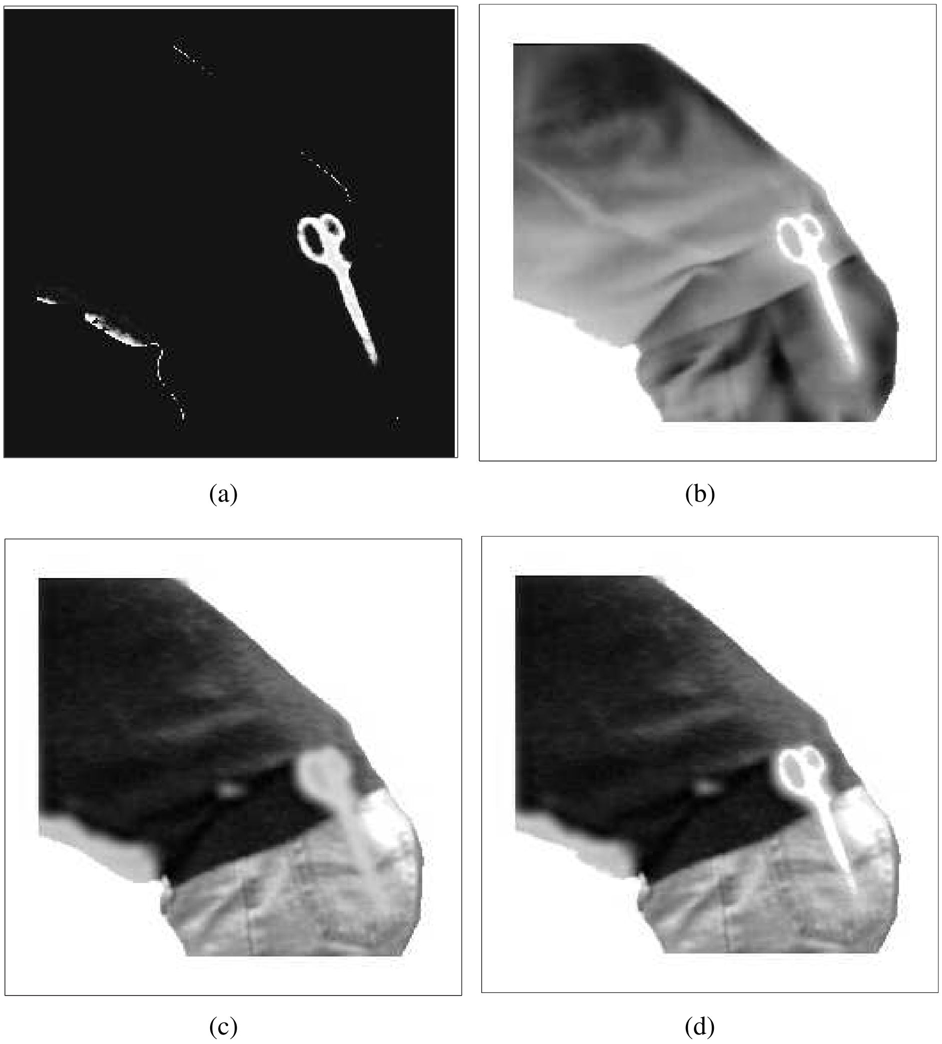


Figure 4.13: The performance of clustering algorithms for IR image of Group A in Figure 4.8(a).

not implement the shaded block in Figure 4.4(b), which may involve two or more long-wavelength sensors for a decision-level fusion. Following the procedure in Figure 4.4(b), we investigated detecting concealed weapons from the IR image and creating a composite image with visual information for an operation or avoiding privacy offense. As far as the second scenario is concerned, the idea is to detect the concealed weapon from the IR, MMW image, or their fusion result and embed the weapon region in the visual image. Since the most important information provided by IR or MMW image is the region of the concealed weapon, the other parts will not do any contribution to the specific analysis. The critical issue is the detection of weapon from IR images. If the weapon cannot be identified, it does not make any sense to fuse it with the visual image.

From the above experiments, we find that the multiresolution-based fusion approaches


(a)

(b)

(c)

(d)

Figure 4.14: Enhancement of ROI: (a) clustered result on the ROI of IR image; (b) enhanced IR image; (c) mosaic result with original IR image; and (d) mosaic result with enhanced IR image.

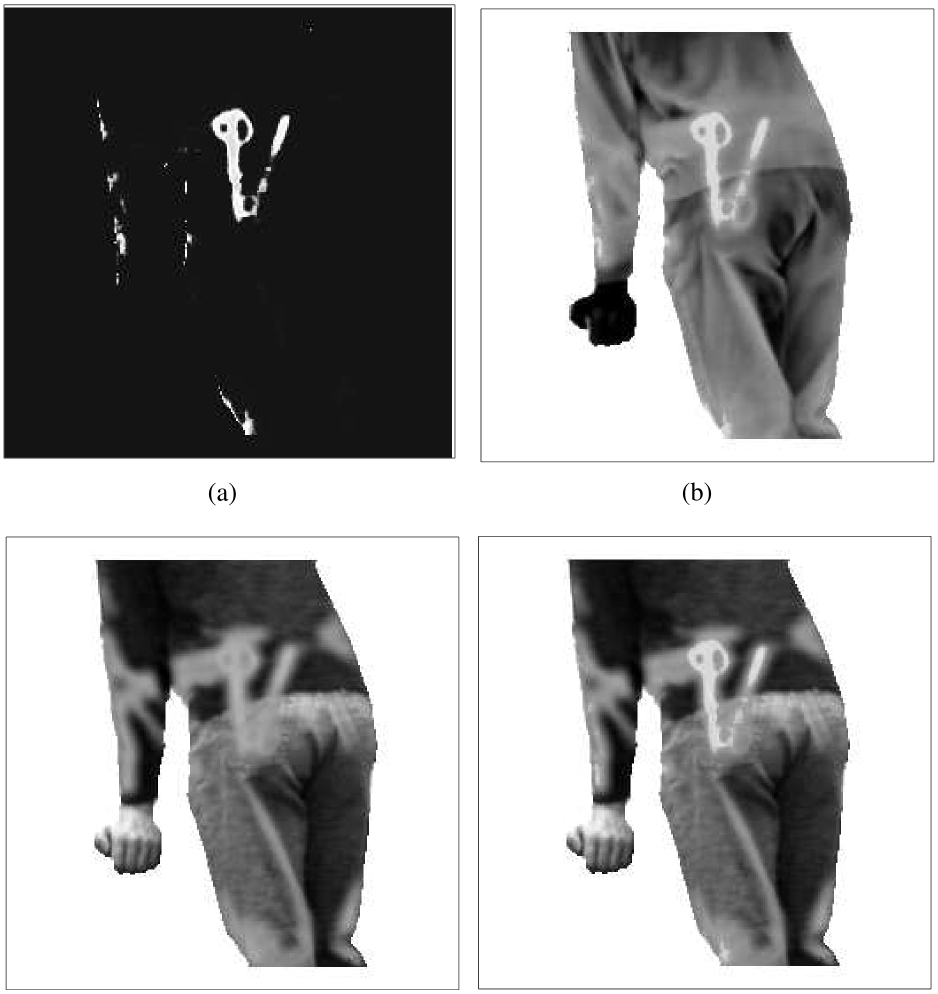
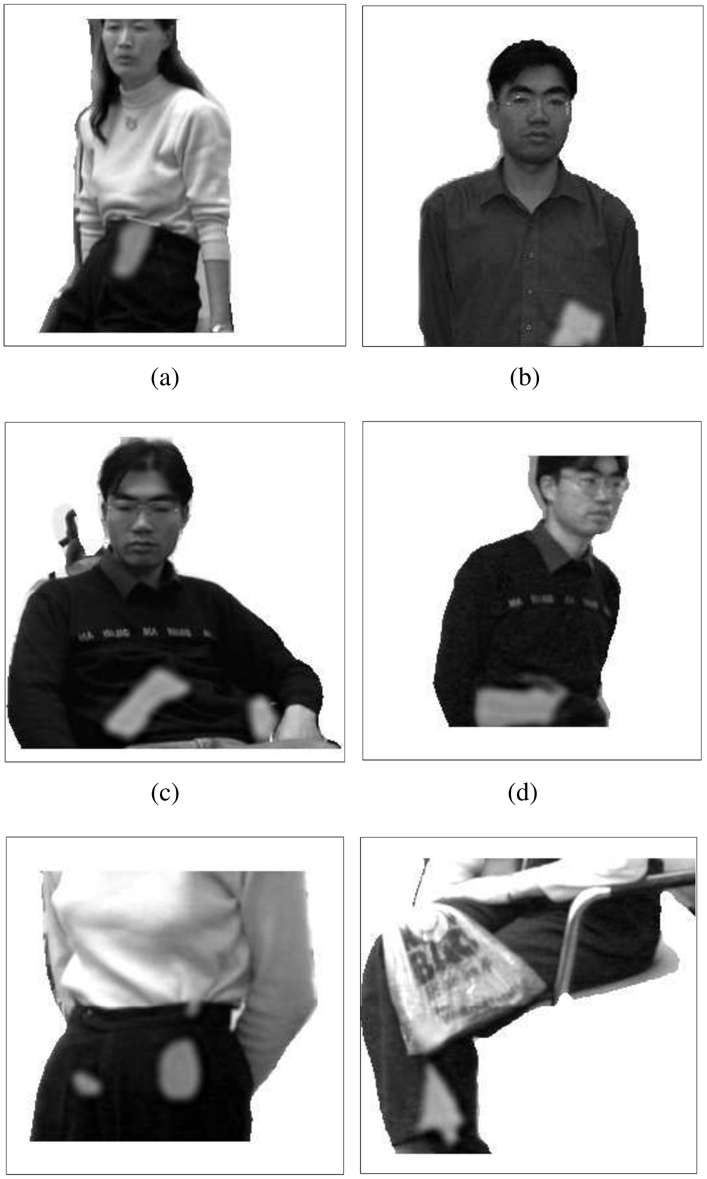


Figure 4.15: Enhancement of ROI: (a) clustered result on the ROI of IR image; (b) enhanced IR image; (c) mosaic result with original IR image; and (d) mosaic result with enhanced IR image.



(a) (b)

(c) (d)

(e) (f)

Figure 4.16: Experimental results achieved by applying the third multiresolution mosaic scheme.

do not always generate a good result. This is due to the variations in image formation and intensity map. Furthermore, the fusion operation degraded the quality of the results due to the integration of useless information. The face is hard to identify in the fused image although the concealed weapon region is highlighted to some extent. Quantitative evaluation of image fusion results is performed by comparing with a reference. The metrics for comparison of two images like root mean square error, correlation, and signal to noise ratio are employed [11], but these values do not assure the fidelity of the fused image. The quality of the fused image can be tested by further processing, such as face recognition or weapon template matching, if applicable. A better fusion result should facilitate the further processing. With the mosaic technique, the visual image's quality can be preserved. The objective assessment of the results is accomplished by using the accuracy and reliability measurements once the threshold value is selected.

The advantage of using fuzzy k-means algorithm is that the clustered pixels are accompanied with a membership value ranging from 0 to 1, which provides additional information, i.e. to what extent we can trust the results. As shown in the experiment, the membership map can also be used for enhancement of the detected ROI region. The clustering does introduce the false alarm due to the "noise" in the IR image, which may come from the background. The detection of foreground object is not a difficult problem to solve. One solution is to use the technique for background subtraction in [105]. A camera calibration procedure is given in [49]. Thus, the processing can be focused on the derived target object. An example is shown in Figure 4.16(c), where the region under that person's left arm should be the backrest of the chair. However, this region is also detected and embedded in the corresponding visual image (see Figure 4.16(c)), although this does not affect the subjective evaluation. Besides, the IR imager distinguishes the weapon from the other part of the body based on the temperature distribution. The bottom edge of the pants and the shirt or some other parts may have the same temperature as the weapon. The clustering algorithm may cluster those regions as concealed weapon too. This does happen to most of the images in Fig. 4.8. The IR imager has its limitation and does not assure a hundred percent detection. Therefore, in order to improve the probability of detection, other image sensor like millimeter wave imager or ultrasound imager can be employed to decrease the uncertainty with more complementary information. The study on probability of detection (POD) should be carried out and higher level fusion can be considered. As suggested by Currie et al. [106], the best use of MMW is for CWD at short range while the best use of IR imager is for wide-area surveillance under poor lighting conditions. The algorithm proposed in this chapter can be applied to visual MMW image pairs.

So far as the multiresolution image mosaic is concerned, one observation is that the process with a larger decomposition level degrades the mosaic results. This is not always true and largely depends on the size of the region (image) to be embedded. When the region is relatively small, as the test images in this chapter, at a lower resolution, the image components will be blurred by the weighted summation with the Gaussian components of the binary mask image. This also happens to the edge-based mask image. Nevertheless, a lower-level decomposition is good for improving the computational efficiency.

In a practical CWD application, the first question to answer is whether there is a weapon concealed underneath clothing. This is a critical step of the whole CWD process. However, this topic is beyond the range of this chapter. This issue is related to the capability of a specific detecting technique, which can be represented by a "probability of detection" (POD) value. As illustrated in Figure 4.4, only when the weapon is present, the combination with a visual image is reasonable. By the way, it is worth mentioning that color-based approach may also be able to provide an effective solution, although this chapter does not cover this topic.

4.6 Conclusion

In this chapter, a scheme based on multiresolution mosaic for concealed weapon detection is presented. The synthesized images are objectively assessed. Certain criteria should be set up to estimate the quality of image mosaic process. The technique will enhance the portal detection for potential threats at the airport or other sensitive locations. The procedure includes two steps: 1) weapon region detection from the IR image and 2) the ROI (detected weapon) mosaic on the visual image. This strategy clarifies the task for each stage, i.e. what to detect and how to combine the results. The multiresolution mosaic technique provides a way to combine two images seamlessly. In the synthesized image, the fidelity of the visual image is preserved well while the concealed weapon is highlighted. An enhancement of the ROI will further facilitate the process. The disadvantage is that the detection algorithm may introduce false positive or false negative error. This is partly due to the limitation of the IR image sensor itself. To improve the probability of detection, information fusion with other image sensors like a millimeter wave imager remains the work for the future.

Chapter 5

The Use of Phase Congruency for Reference-based Assessment

5.1 Introduction

As demonstrated in Chapter 3 and 4, the image fusion can be implemented with different (multiresolution) algorithms. In order to identify the performance of the algorithm for a specific application, there raises the problem of fusion performance assessment. The assessment of a fused image can be carried out by comparing it with a reference image, which is assumed to be perfect for a specific application. Usually, such comparison is implemented on pixel-based operations, like mean square error (MSE) or root mean square error (RMSE). However, such operations' performance is questionable because the same MSE or RMSE value does not always assure a comparable image similarity under different distortion to perceptually significant features [107].

According to Wilson et al. [108], there are three major methods to compare images: human perception, objective measures based on theoretical models, and subjective measures defined mathematically. In their publication [108], the authors used a distance measure of two sets of pixels to characterize the numerical difference between two images. Metrics for binary image (Δ_b) and gray-scale image (Δ_g) comparison were developed respectively. Lees et al. used a phase-only approach for printed circuit board (PCB) inspection [109]. The phase-only imaging has the advantages of being light intensity invariant, insensitive to illumination gradients, and tolerant to misregistration. Therefore, it is suitable for the application of PCB inspection in electronic packaging industry. Lorenzetoo and Kovesi developed this work by computing the phase difference for distortion measurement [110]. The superority of the proposed algorithm is its ability to discriminate translation from distortion.

In practical applications, some post-processing largely depends on the availability of image features. Operations, like classification, segmentation, and quantification, will be carried out in a feature space. Therefore, the availability of image features plays an important role for further analysis. This chapter proposes two new methods to quantitatively assess image by employing a phase congruency measurement proposed by Kovesi [3, 111]. The phase congruency measurement provides an absolute measure of image features like step edges, lines and Mach bands; therefore, it is viewing condition-independent and irrespective of image illumination and magnification. In the first method, a local cross-correlation of the phase congruence map is calculated for the fused image and reference image. The averaged value provides a quantitative assessment of the overall image similarity. In the second method, a modified image structural similarity measurement (SSIM) is implemented by replacing the structure component $s(a, b)$ with the correlation between the phase congruency maps. Experiments are carried out on standard images and fused images.

This chapter deals with the reference-based methods for fusion performance assessment and next chapter will move to the evaluation without a reference. The rest of this chapter is organized as follows: the available solutions for image comparison are presented in section (5.2). In section (5.3), a description of the phase congruency measurement is provided. Then, two reference-based metrics are proposed in section (5.4). Experimental results are demonstrated in section (5.5). Discussion and conclusion can be found in section (5.6) and (5.7) respectively.

5.2 Typical Solutions

There are a number of metrics available for image comparison. The commonly used approaches include root mean square error, normalized least square error (NLSE), the peak signal to noise ratio (PSNR), and correlation (CORR). The definition of these metrics are given in equation (5.1)-(5.4), where $I\left(x,y\right)$ and $F\left(x,y\right)$ stand for the reference and input (fused) image respectively and L is the maximum pixel value. The size of the two images is M-by-N. The advantage of the these methods is the simplicity and computational efficiency. However, there is the possibility that images with a similar RMSE value may exhibit a quite different appearance. It is a disadvantage of the RMSE method.

$$RMSE=\sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\left[F\left(x,y\right)-I\left(x,y\right)\right]^2}{MN}} \tag{5.1}$$

$$NLSE=\sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\left[F\left(x,y\right)-I\left(x,y\right)\right]^2}{\sum_{x=1}^{M}\sum_{y=1}^{N}\left[F\left(x,y\right)\right]^2}} \tag{5.2}$$

$$PSNR=10\log_{10}\left(\frac{L^2}{\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\left[F\left(x,y\right)-I\left(x,y\right)\right]^2}\right) \tag{5.3}$$

$$CORR=\frac{2\sum_{x=1}^{M}\sum_{y=1}^{N}F\left(x,y\right)I\left(x,y\right)}{\sum_{x=1}^{M}\sum_{y=1}^{N}R\left(x,y\right)^2+\sum_{x=1}^{M}\sum_{y=1}^{N}I\left(x,y\right)^2} \tag{5.4}$$

Table 5.1: The notation for equation (5.1)-(5.5).

$I(x, y)$	reference image
$F(x, y)$	fused image
L	maximum pixel value
$P_I(g)$	probability of value g of the reference image
$P_F(g)$	probability of value g of the fused image
$h_{IF}(u, v)$	normalized joint histogram of the reference image and he fused image
$h_I(u)$	normalized marginal histogram of the reference image
$h_F(v)$	normalized marginal histogram of the fused image

More sophisticated methods include difference entropy (DE), mutual information (MI), and structural similarity index measure [53, 39]. The difference entropy between two images reflects the difference between the average amount of information they contained. It is defined as:

$$DE = \left| \sum\nolimits_{g=0}^{L-1} P_F\left(g\right) \log_2 P_F\left(g\right) - \sum\nolimits_{g=0}^{L-1} P_I\left(g\right) \log_2 P_I\left(g\right) \right| \tag{5.5}$$

where $P_I\left(g\right)$ and $P_F\left(g\right)$ are the probability of pixel value g for the reference and input image respectively. The mutual information between the input and reference images is defined on the normalized joint gray level histogram $h_{FI}\left(u, v\right)$ and normalized marginal histogram of the two images i.e. $h_F\left(u\right)$ and $h_I\left(v\right)$:

$$MI = \sum\nolimits_{v=1}^{L} \sum\nolimits_{u=1}^{L} h_{FI}\left(u, v\right) \log_2 \frac{h_{FI}\left(u, v\right)}{h_F\left(u\right) h_I\left(v\right)} \tag{5.6}$$

The notation for the equation (5.1) to (5.5) is given in Table 5.1. The SSIM algorithm is presented with the equation (2.7) in Chapter 2.

The purpose of this section is not to provide a full list or review of the available image comparison metrics. New solutions are still being developed and proposed. Only the typical metrics, which have been frequently used to evaluate the performance of image fusion algorithms, are described and used in the following experiments.

5.3 Image Feature from Phase Congruency

5.3.1 The Concept of Phase Congruency

Gradient-based image feature detection and extraction approaches are sensitive to the variations in illumination, blurring, and magnification. The threshold applied needs to be modified appropriately. A model of feature perception named local energy was investigated by Morrone and Owens [112]. This model postulates that features are perceived at points in an image where the Fourier components are maximally in phase. A wide range of feature types give rise to points of high phase congruency. With the evidence that points of maximum phase congruency can be calculated equivalently by searching for peaks in the local energy function, the relation between the phase congruency and local energy is established; that is [113, 114]:

$$PC(x) = \frac{E\left(x\right)}{\sum_n A_n\left(x\right) + \varepsilon} \tag{5.7}$$

$$E\left(x\right) = \sqrt{F^2\left(x\right) + H^2\left(x\right)} \tag{5.8}$$

where $PC\left(x\right)$ is the phase congruency at some location x and $E\left(x\right)$ is the local energy function. A_n represents the amplitude of the n^{th} component in the Fourier series expansion. A very small positive constant ε is added to the denominator in case of small Fourier amplitudes. In the expression of local energy, $F\left(x\right)$ is the signal with its DC component removed and $H\left(x\right)$ is the Hilbert transform of $F\left(x\right)$.

5.3.2 Implementation of Phase Congruency Algorithm with the Logarithmic Gabor Filter

Kovesi proposed a scheme to calculate phase congruency using logarithmic Gabor wavelets [3, 113, 115, 114, 116], which allow arbitrarily large bandwidth filters to be constructed while still maintaining a zero DC component in the even-symmetric filter. However, when there are few high frequency components in the signal, the frequency spread is reduced. The phase congruency will be one everywhere if the signal is a pure sine wave. To counter this problem, a weighting function $W\left(x\right)$ is constructed to devalue phase congruency at locations where the spread of filter response is narrow. To further enhance the calculation, Kovesi brought in a more sensitive phase deviation $\Delta\Phi(x)$ to define the phase congruency, i.e.

$$\Delta\Phi\left(x\right) = \cos\left(\phi_n(x) - \bar{\phi}(x)\right) - |\sin\left(\phi_n(x) - \bar{\phi}(x)\right)| \tag{5.9}$$

where $\phi_n(x)$ and $\bar{\phi}(x)$ are the phase angle and overall mean respectively. The equation of the new phase congruency measure now becomes [113, 114]:

$$PC(x) = \frac{\sum_n W(x) \left\lfloor A_n(x)\Delta\Phi_n(x) - T \right\rfloor}{\sum_n A_n(x) + \varepsilon} \tag{5.10}$$

where $\lfloor\rfloor$ denotes that the enclosed quantity is not permitted to be negative. Here, T is an estimated compensation for the noise influence. The detailed implementation of the algorithm is given in Appendix A.

To extend the algorithm to images, the one-dimensional analysis is applied to several orientations and the results are combined. The 2D phase congruency can be expressed as [113, 114]:

$$PC\left(x\right) = \frac{\sum_o \sum_n W_o(x) \left\lfloor A_{no}(x) \Delta\Phi_{no}(x) - T_o \right\rfloor}{\sum_o \sum_n A_{no}(x) + \varepsilon} \tag{5.11}$$

where o denotes the index over orientation. The noise compensation T_o is performed in each orientation independently. By simply applying the Gaussian spreading function across the filter perpendicular to its orientation, the one-dimensional Gabor filter can be extended into two dimensions. The orientation space can be quantified using a step size of $\pi/6$, which results in six different orientations. For an extensive discussion of the underlying theory, references [114, 113] are recommended.

A flowchart that describes the procedure of computing the image phase congruency map of image "Einstein" is given in Figure 5.1. Only one orientation is shown and the final result (phase congruency map) is obtained by summing the results along all the predefined orientations.

5.4 Reference-based Assessment for Image Fusion

5.4.1 Image Similarity Measurement

The comparison of images can be carried out by comparing their corresponding phase congruency features. It is appropriate to evaluate the space-variant features locally and combine them together [40, 53]. As suggested by Wang et al. [40], a sliding window of size 5-by-5 is moved from top-left to bottom-right of the image. Thus, at each location, we have a sub-block image. A straightforward solution is to compare (sub-block) images with their phase congruence maps based on the cross-correlation directly [117]. However, in a phase congruency map, the sub-block window could be blank, i.e. all the feature points are zero at the locations without any features or with a local energy value less than the threshold value T_o in equation (5.11). The immediate result cannot be obtained due

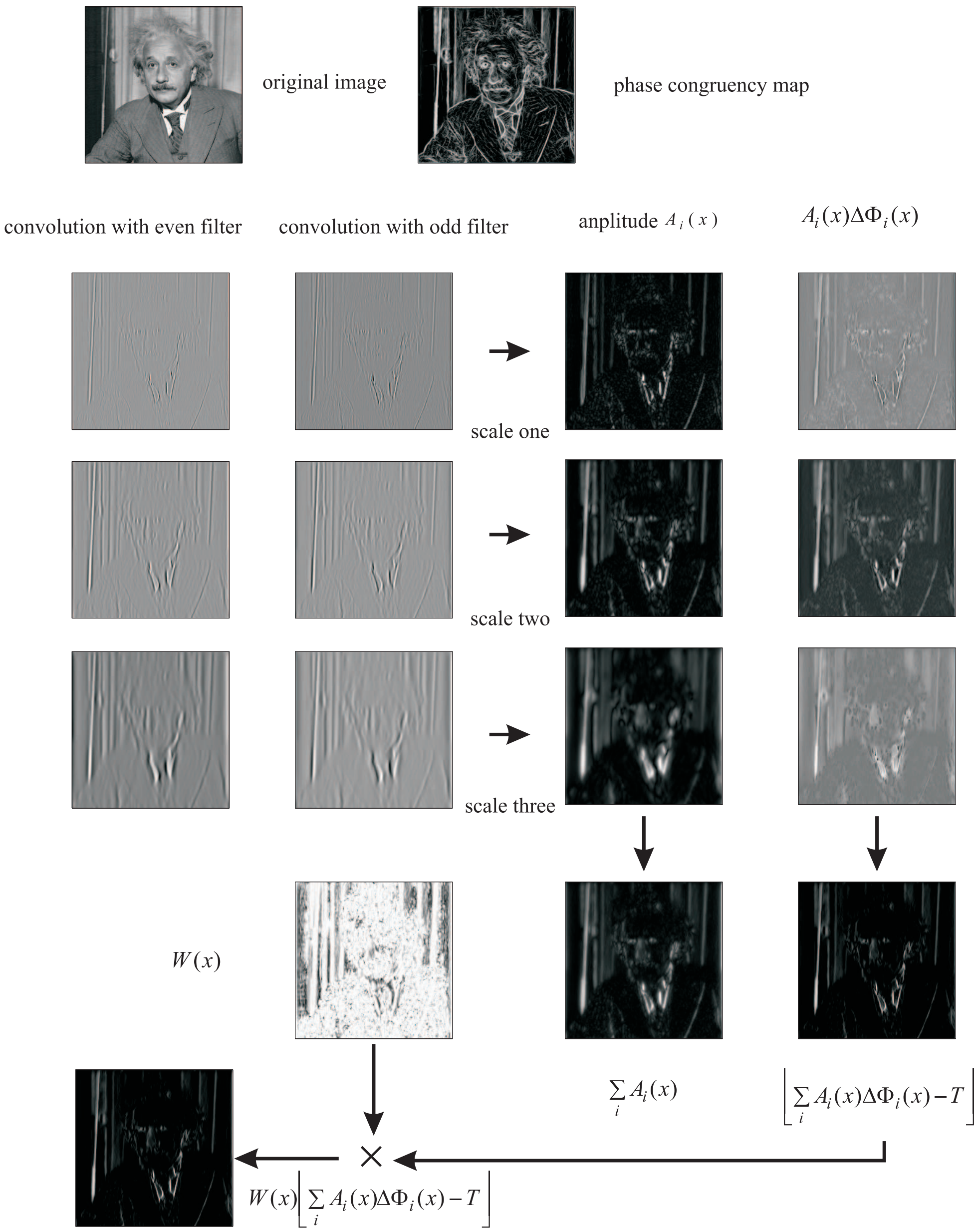

original image
phase congruency map
convolution with even filter
convolution with odd filter
anplitude $A_i(x)$
$A_i(x)\Delta\Phi_i(x)$
scale one
scale two
scale three
$W(x)$
$\sum_i A_i(x)$
$\left\lfloor \sum_i A_i(x)\Delta\Phi_i(x) - T \right\rfloor$
$\times$
$W(x)\left\lfloor \sum_i A_i(x)\Delta\Phi_i(x) - T \right\rfloor$

Figure 5.1: The calculation of phase congruency map (one orientation is presented).

to the zero in the denominator of cross-correlation's expression. Figure 5.2 indicates an alternative procedure to calculate the local cross-correlation.

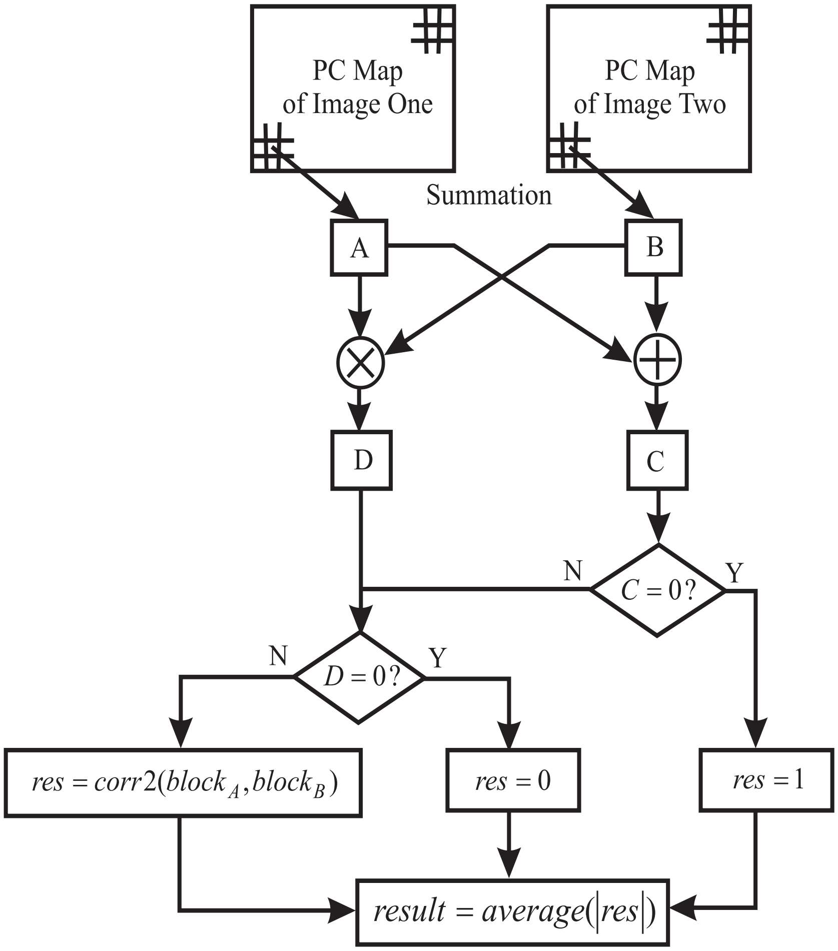


Figure 5.2: The P_{ref} metric for reference-based evaluation.

First, the pixels in the sub-block window from the two images are summed respectively to get the results A and B. The summation and product of these two values are C and

D respectively. If the summation C appears to be zero, which means both A and B are zero and these two image blocks are totally matched. Therefore, the corresponding cross-correlation value should be set to 1. When the block is different, i.e. $C \neq 0$ and $D = 0$, the cross-correlation value is set to 0. Otherwise the value is computed by the zero-mean normalized cross-correlation (ZNCC) [117] in equation (5.12):

$$ZNCC = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \left(I_1\left(x,y\right) - \bar{I}_1\right)\left(I_2\left(x,y\right) - \bar{I}_2\right)}{\sqrt{\sum_{x=1}^{M} \sum_{y=1}^{N} \left(I_1\left(x,y\right) - \bar{I}_1\right)^2 \cdot \sum_{x=1}^{M} \sum_{y=1}^{N} \left(I_2\left(x,y\right) - \bar{I}_2\right)^2}} \tag{5.12}$$

where $\bar{I}_1$ and $\bar{I}_2$ are the average value of the two $M \times N$ images (sub-images) $I_1(x,y)$ and $I_2(x,y)$ respectively. The final result in Figure 5.2 gives the metric P_{ref}.

5.4.2 A Modified SSIM Scheme

The second consideration is to incorporate the phase congruency measurement into the structural similarity framework proposed by Wang et al. [53]. The major problem with SSIM is that it fails to measure severely blurred image [118]. A discussion of this problem is presented in Appendix B. Chen et al. proposed edge-based structure measurement to replace the $s(a,b)$ component in equation (2.7) [118]. The edge direction histogram was used to compare the edge information. Instead of using the edge direction vector, we employ the phase congruency measurement. The feature-based SSIM (FSSIM) becomes:

$$FSSIM\left(a,b\right) = \left[l\left(a,b\right)\right]^{\alpha}\left[c\left(a,b\right)\right]^{\beta}\left[f\left(a_p,b_p\right)\right]^{\gamma} \tag{5.13}$$

Herein, the feature component is defined as the correlation of two phase congruency maps:

$$f\left(a_p, b_p\right) = \frac{\sigma_{a_p b_p} + \varepsilon}{\sigma_{a_p} \sigma_{b_p} + \varepsilon} \tag{5.14}$$

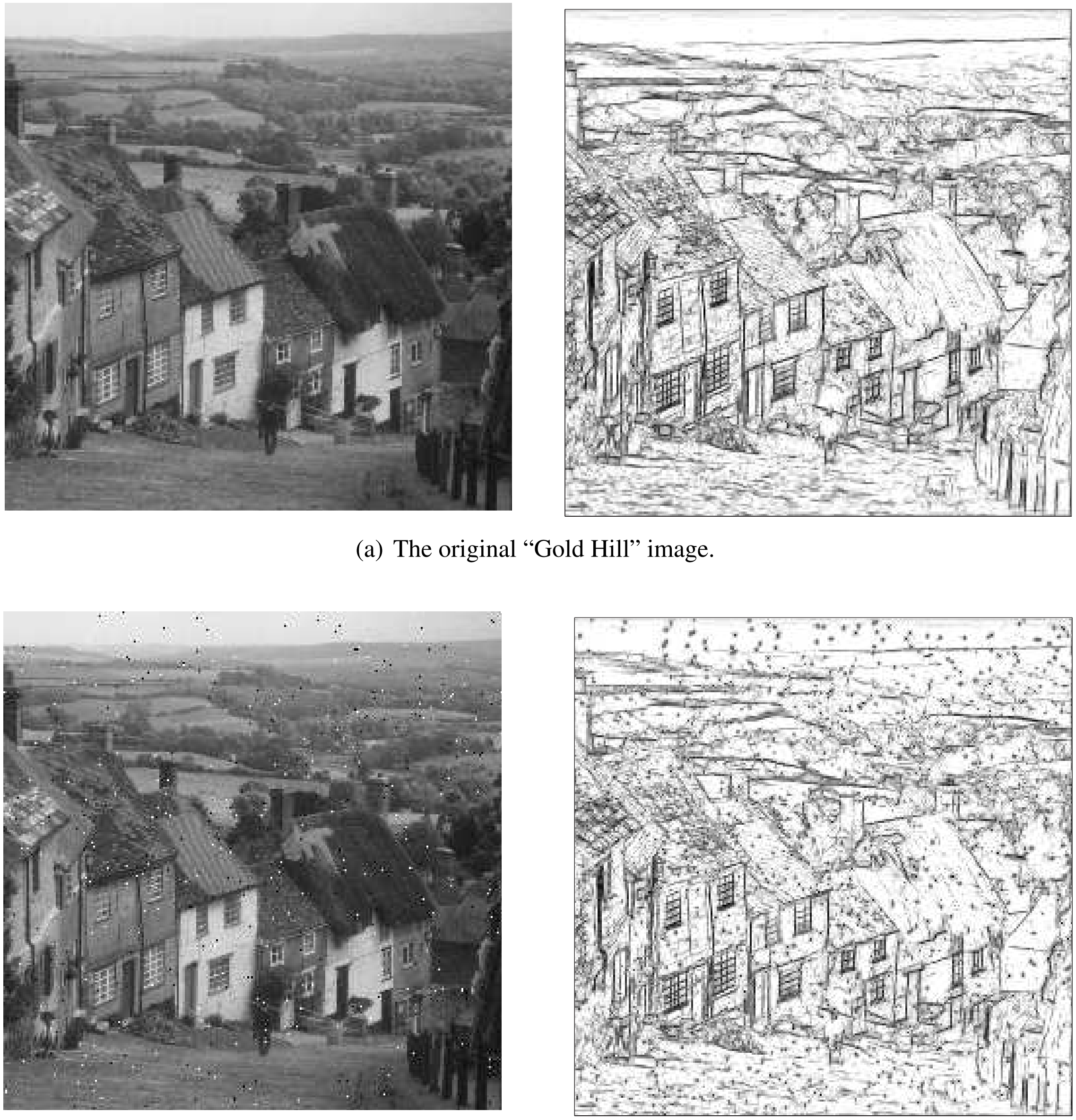
where a small constant is added to avoid the denominator being zero. Similarly, the covariance and standard deviation of two phase congruence maps a_p and b_p are $\sigma_{a_p b_p}$, σ_{a_p} and σ_{b_p} respectively. The equation (5.14) is actually an expression for calculating the correlation. We treat the three factors equally and the parameter α, β, and γ are all set to one in the following experiments.

5.5 Experimental Results

5.5.1 Experiments for Image Comparison

The first part of the experiments is carried out on a set of images as shown in Figure 5.3 to Figure 5.10. These images were prepared to have an identical root mean square error; however, the appearance is totally different. The original image is contaminated by salt-pepper noise, Gaussian noise, speck noise, mean shifting, contrast stretching, blurring operation, and JPEG compressing respectively. Besides the RMSE, a group of metrics are also computed for the comparison: normalized least-square error, peak signal-to-noise ratio, correlation, difference entropy, mutual information, structural similarity measure, and the two proposed methods. The numerical results are listed in Table 5.2 and 5.3.

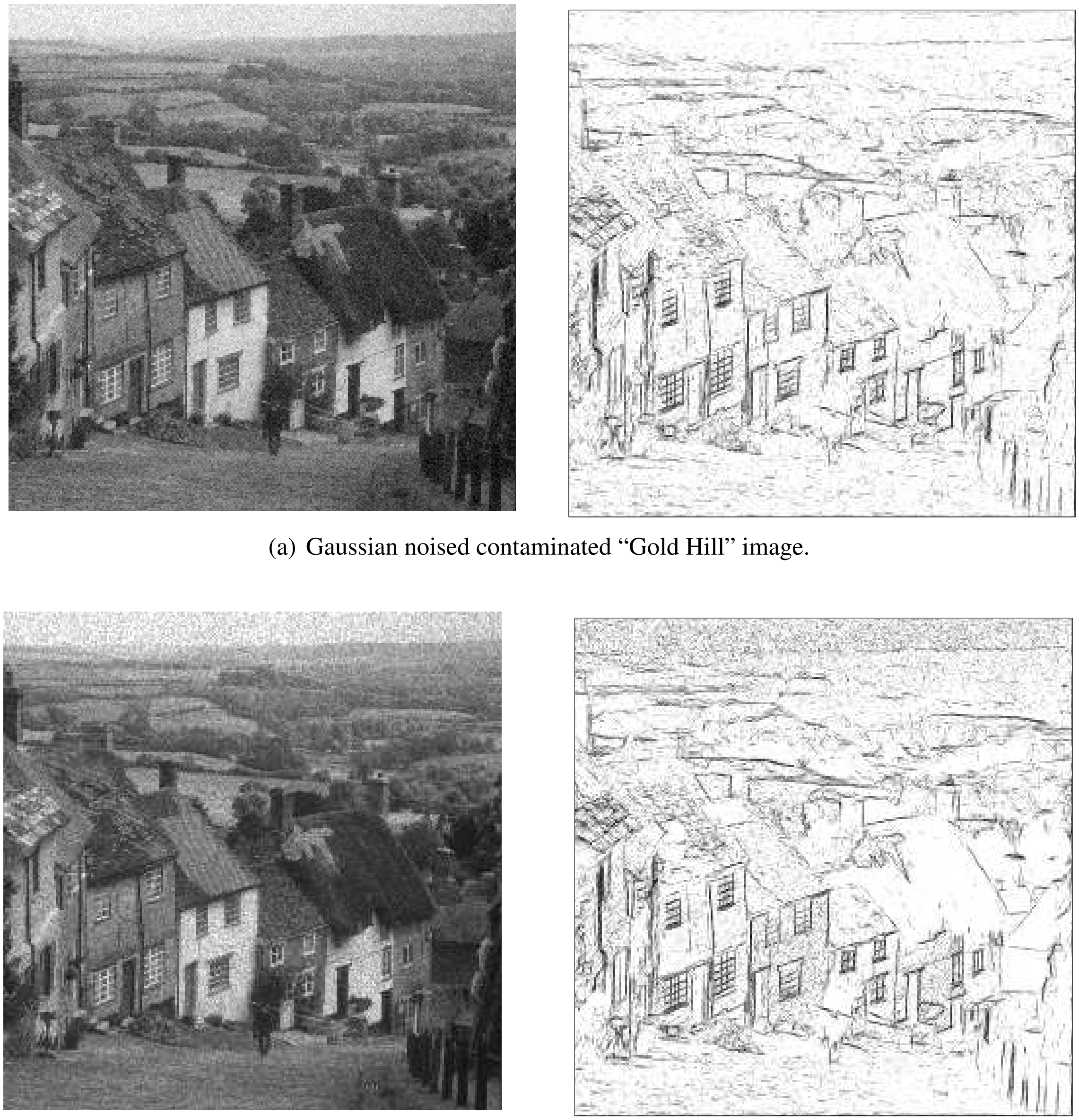
In the numerical results presented in Table 5.2, it is hardly to identify the difference from RMSE and NLSE. PSNR does distinguish the different distortions. However, it obtains a similar value for Gaussian noise and blurring operation. Although a larger value of MI indicates a good resemblance, it is not a normalized parameter and we still need a reference to compare. Similarly, although the zero of DE indicates a perfect match of two images, a reference value is necessary in most cases. Table 5.3 indicates that the affects



(a) The original “Gold Hill” image.

(b) Salt-pepper noise contaminated “Gold Hill” image.

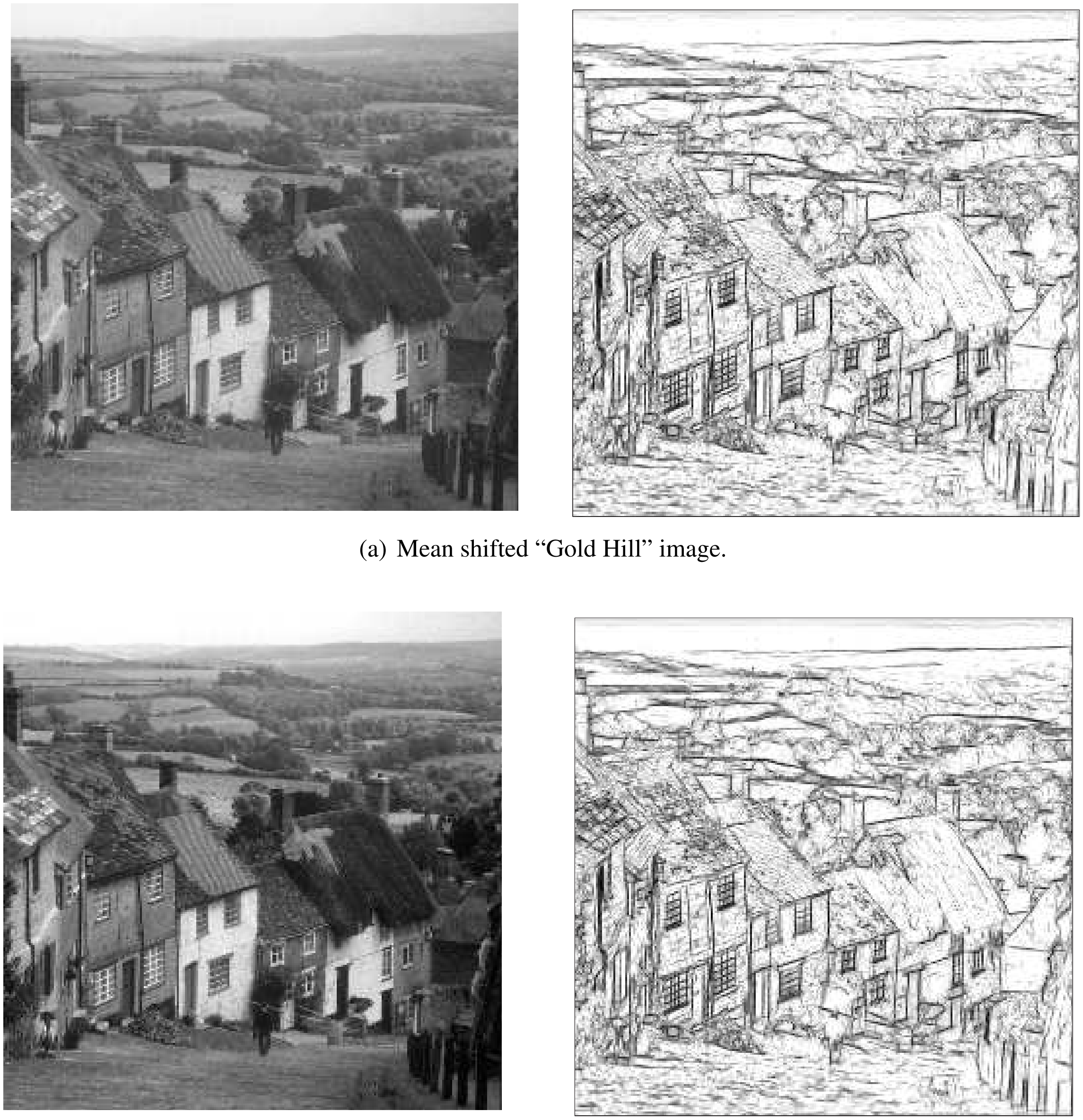
Figure 5.3: The “Gold Hill” image (left) and its phase congruency map (right).



(a) Gaussian noised contaminated “Gold Hill” image.

(b) Speck noise contaminated “Gold Hill” image.

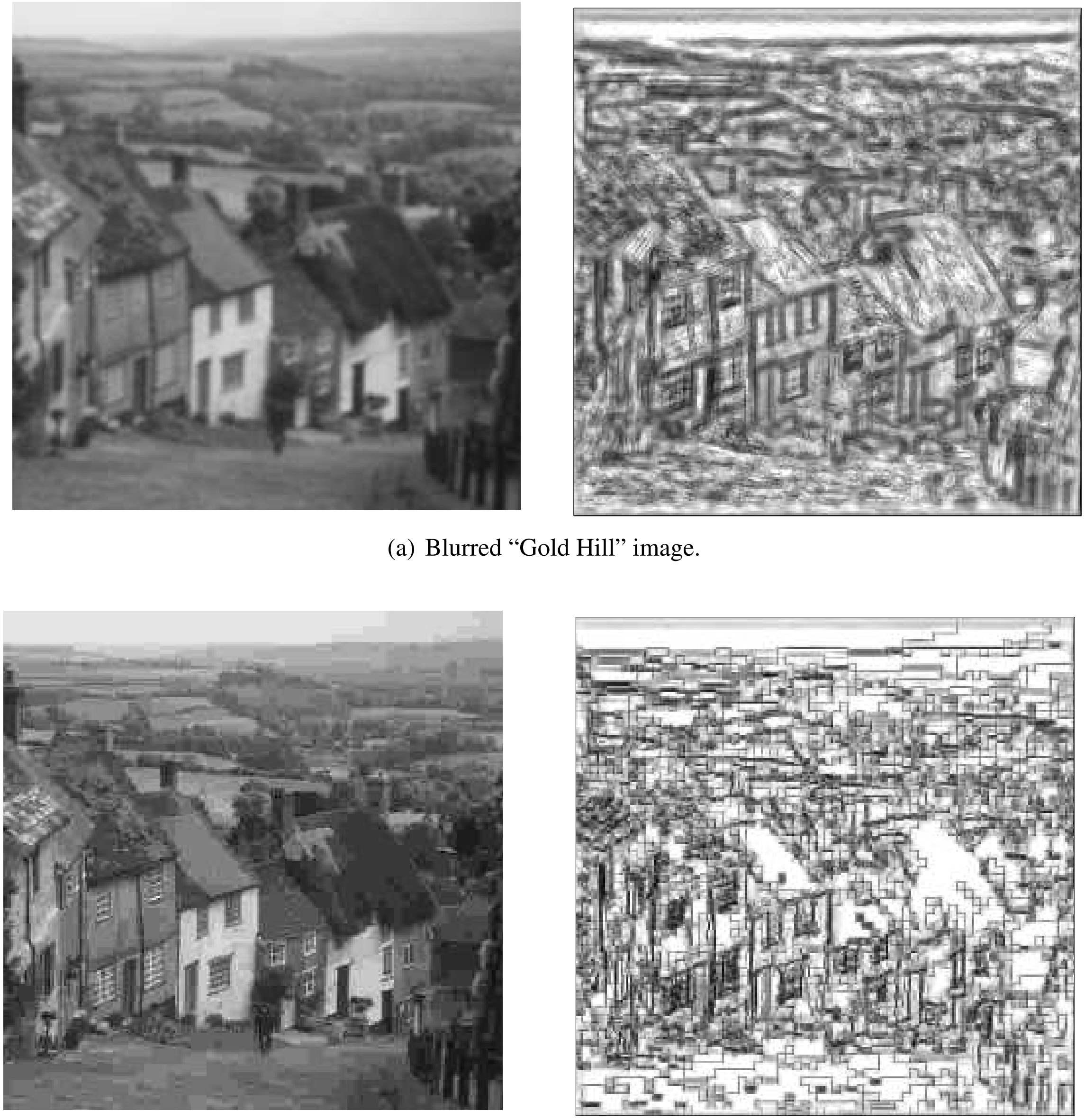
Figure 5.4: The “Gold Hill” image (left) and its phase congruency map (right).



(a) Mean shifted “Gold Hill” image.

(b) Contrast stretch “Gold Hill” image.

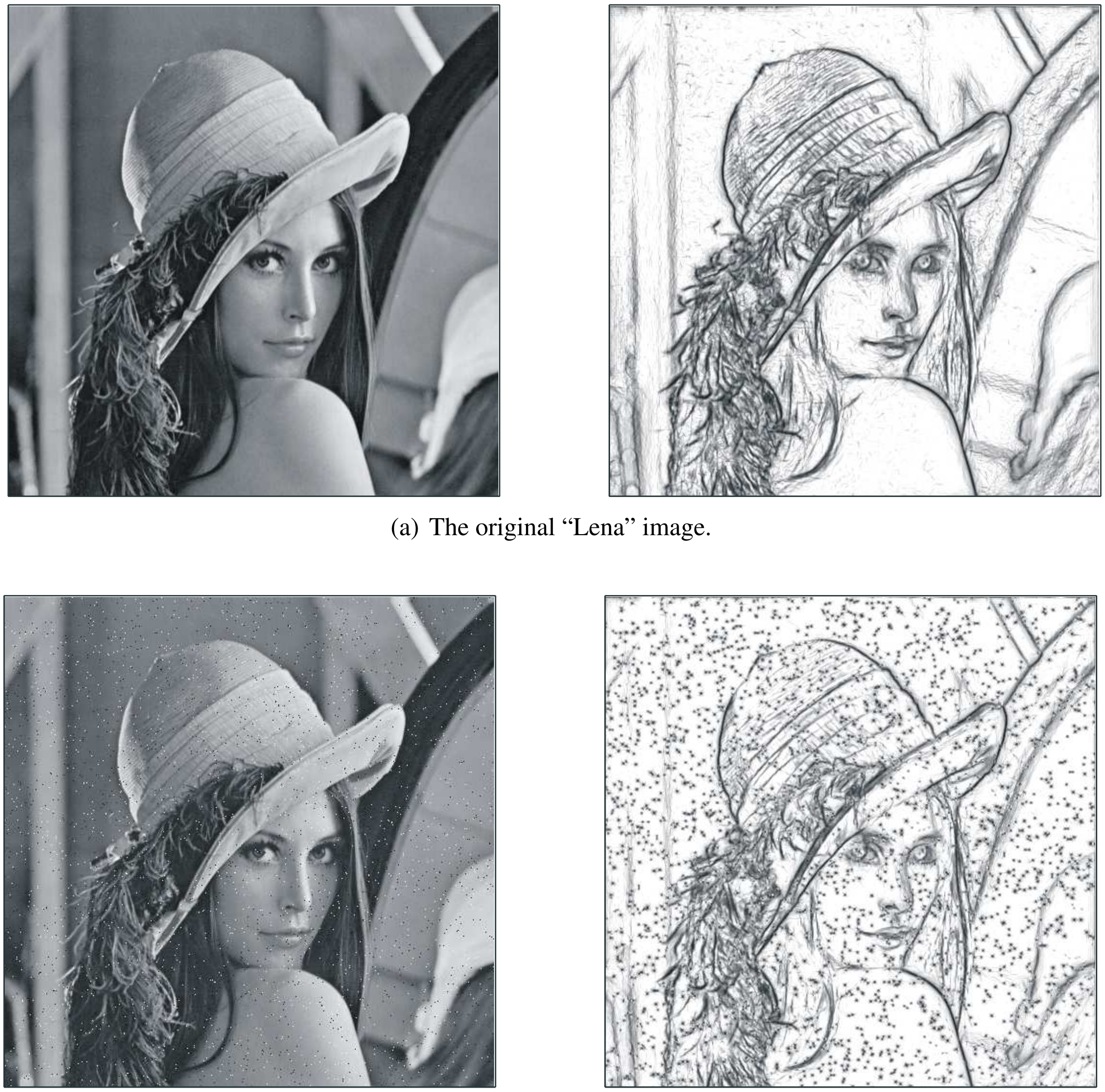
Figure 5.5: The “Gold Hill” image (left) and its phase congruency map (right).



(a) Blurred “Gold Hill” image.

(b) JPEG compressed “Gold Hill” image.

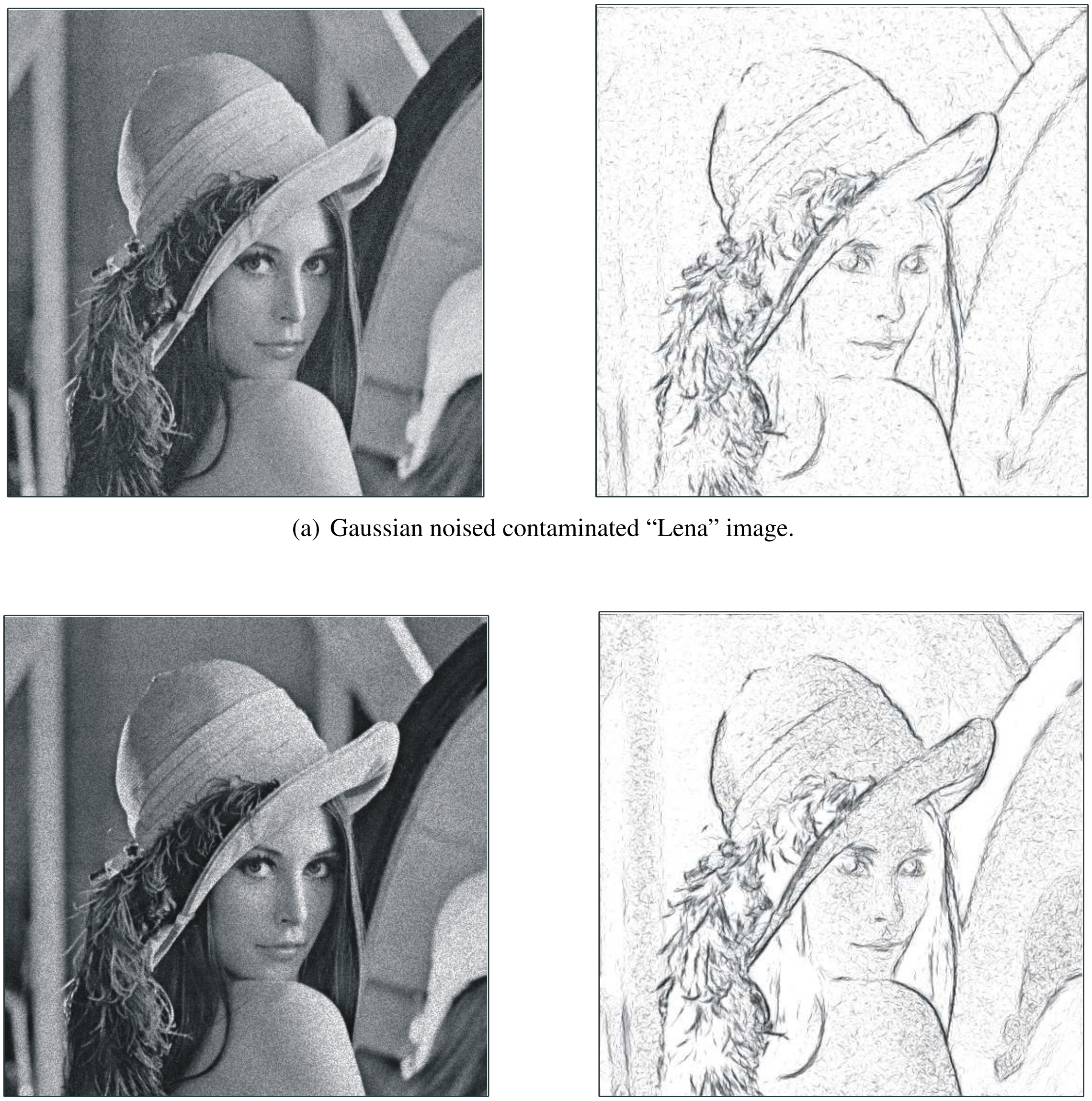
Figure 5.6: The “Gold Hill” image (left) and its phase congruency map (right).



(a) The original “Lena” image.

(b) Salt-pepper noise contaminated “Lena” image.

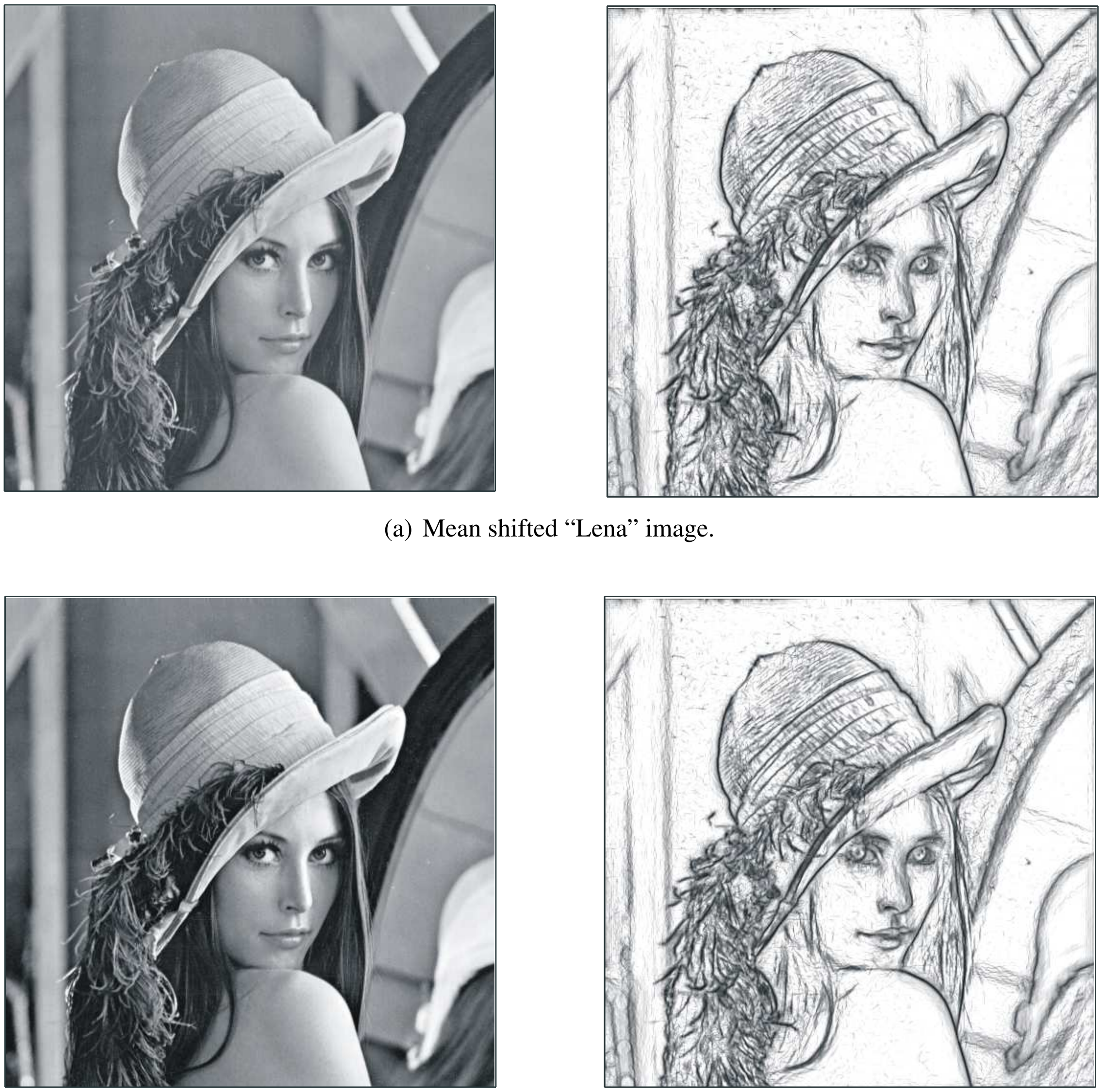
Figure 5.7: The “Lena” image (left) and its phase congruency map (right).



(a) Gaussian noised contaminated “Lena” image.

(b) Speck noise contaminated “Lena” image.

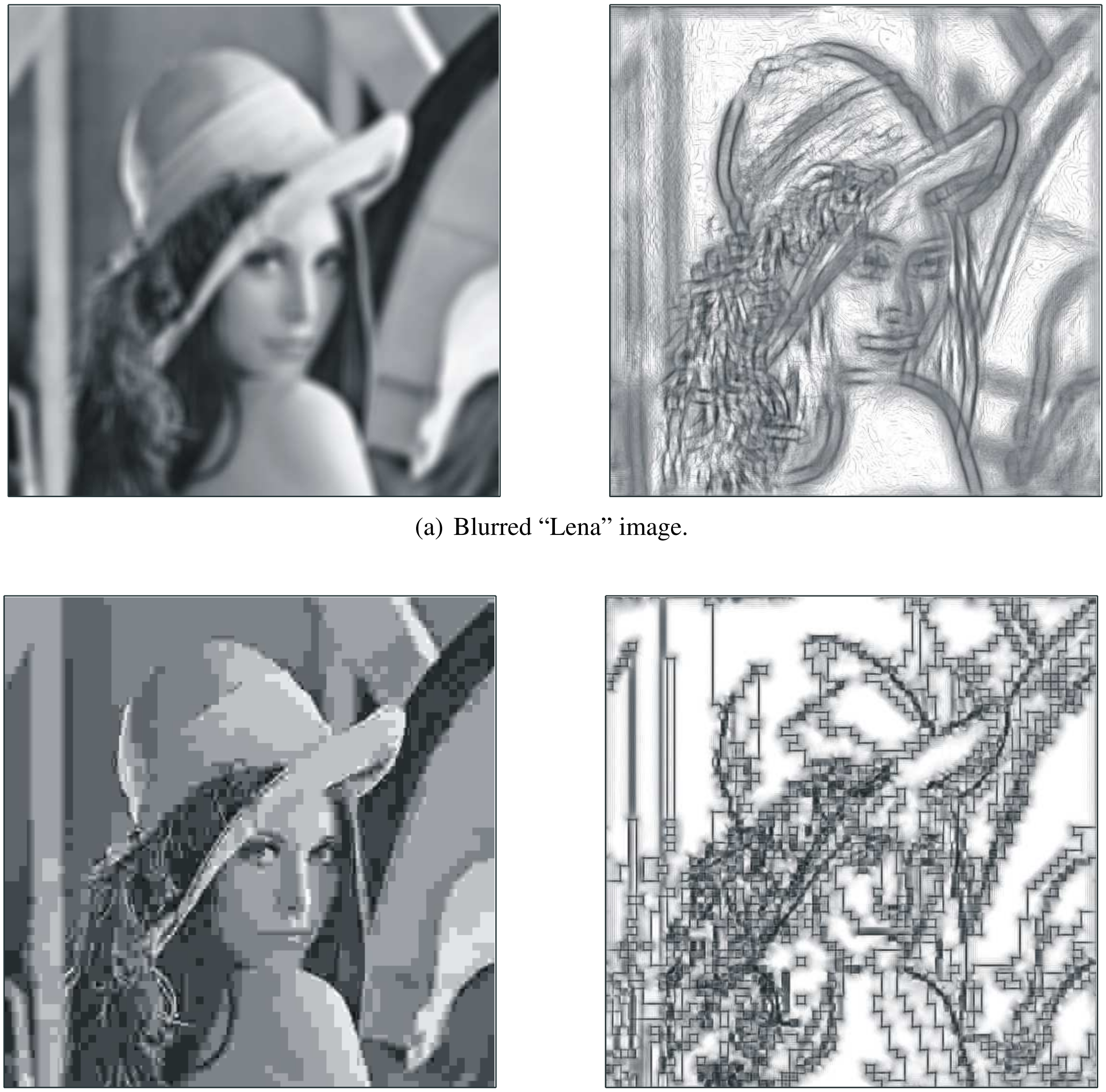
Figure 5.8: The “Lena” image (left) and its phase congruency map (right).



(a) Mean shifted “Lena” image.

(b) Contrast stretch “Lena” image.

Figure 5.9: The “Lena” image (left) and its phase congruency map (right).



(a) Blurred “Lena” image.

(b) JPEG compressed “Lena” image.

Figure 5.10: The “Lena” image (left) and its phase congruency map (right).

of Gaussian noise and Salt-pepper noise are quite similar in terms of PSNR. RMSE and NLSE.

Table 5.2: Experimental results on image comparison (Gold Hill).

Gold Hill Image	(b)	(c)	(d)	(e)	(f)	(g)	(h)
RMSE	10.9640	11.0060	11.0200	11.0000	10.9950	11.0430	10.8390
NLSE	8.9483	8.9823	8.9935	8.9776	8.9739	9.0123	8.8458
PSNR	25.3140	25.4100	24.5020	Inf	38.6860	25.4150	24.9880
CORR	0.9960	0.9960	0.996	0.9963	0.9961	0.9959	0.9961
DE	0.0141	0.1163	0.0754	0.0000	0.0770	0.0934	1.0005
MI	3.4111	1.4347	1.6964	3.9347	3.4654	1.5601	1.4656
SSIM	0.8643	0.6556	0.7032	0.9927	0.9698	0.6671	0.6824
P_{ref}	0.7569	0.5079	0.5311	1.0000	0.9948	0.3478	0.3795
FSSIM	0.8127	0.5543	0.6039	0.9927	0.9658	0.2525	0.3291

Table 5.3: Experimental results on image comparison (Lena).

Lena Image	(b)	(c)	(d)	(e)	(f)	(g)	(h)
RMSE	15.0123	15.006	14.9916	15.00	15.0031	14.9713	14.6668
NLSE	0.1141	0.1141	0.1140	0.1140	0.1140	0.1138	0.1115
PSNR	22.7157	22.8299	23.4438	40.7125	27.4375	22.1800	22.5848
CORR	0.9531	0.9542	0.9543	1.0000	1.0000	0.9510	0.9537
DE	-0.0163	-0.1953	-0.1412	0.0000	0.0067	0.1402	3.1991
MI	3.3484	1.1160	1.4042	3.4120	3.5453	1.4740	1.2560
SSIM	0.7227	0.4508	0.5009	0.9890	0.9494	0.6880	0.6709
P_{ref}	0.4734	0.3781	0.3868	1.0000	0.9904	0.2928	0.2815
FSSIM	0.5357	0.3109	0.3552	0.9890	0.9463	0.0742	0.1721

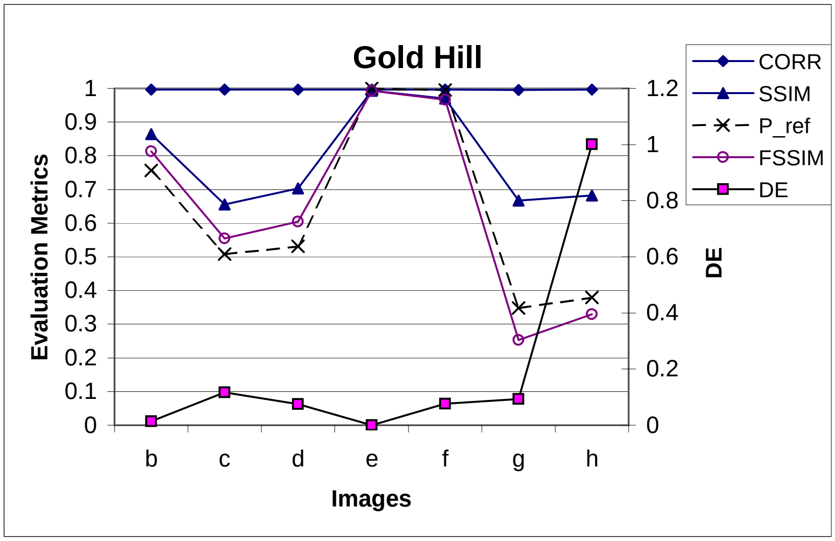
The results obtained with CORR, DE, SSIM, P_{ref}, and FSSIM are plotted in Figure 5.11(a) and 5.11(b). It is hard to find the difference from CORR while the SSIM, P_{ref}, and FSSIM metric demonstrate a quite similar trend.

Mean-shifted images in Figure 5.5(a) and 5.9(a) are obtained by adding the image with a constant value. However, if the mean-shifted image is displayed or manipulated in a different way, the conclusion may be different. The first way is to map the maximum and minimum value to 0 and 255. This does not change the appearance of the original image. The second is setting the pixel value to 255 if it exceeds 255. If the image is processed as the latter one, the operation is not the mean shifting any more and the image appears to be degraded as in Wang and Bovik's publication [40]. To avoid such confusion, we show the mean-shifted image with the first method, i.e. there is no distortion caused by image depth limit. The image appears to be exact the same as its origin. Therefore, such operation does not alter or destroy the structural information conveyed by an image. As a result, the P_{ref} metric reaches it maximum value "one" while the DE value is zero. The SSIM and FSSIM metric give a value close to one.

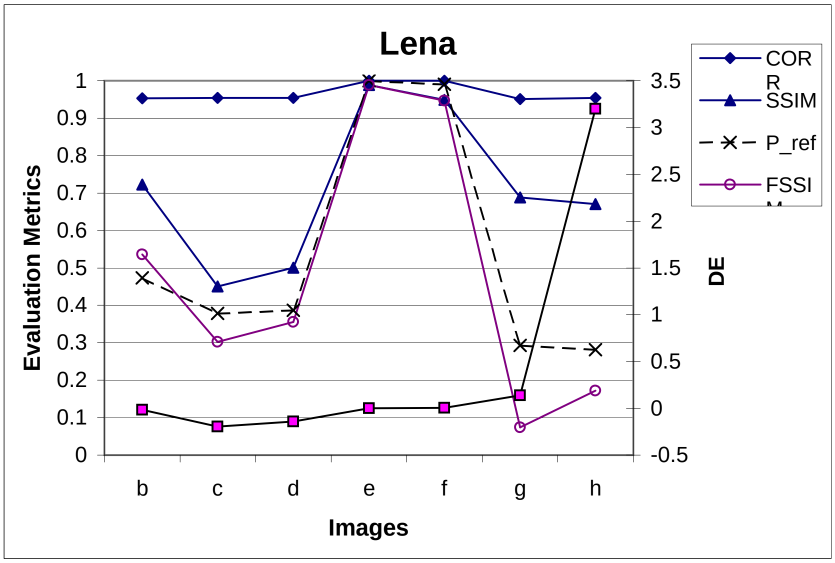
Secondly, we list the standard deviation of the results obtained by SSIM, FSSIM, and P_{ref} in Table 5.4. A larger standard deviation value indicates the difference between these values is make easier to identified.

Table 5.4: The standard deviation of the assessment results for image "gold hill" and "Lena".

	SSIM	P_{ref}	FSSIM
Gold Hill	0.1479	0.2743	0.2932
Lena	0.2034	0.3153	0.3613



(a) Chart for Gold Hill.



(b) Chart for Lena.

Figure 5.11: The chart for image comparison.

5.5.2 Experiments for Fusion Assessment

In the second part of the experiments, the proposed metrics are used to assess the fusion results obtained by different multiresolution algorithms. A group of images considered here consist of multi-focus images from a digital camera and the corresponding full-focus reference images as shown in Figure 5.12. The multi-focus images are fused by using the following algorithms: Laplacian pyramid, gradient pyramid, ratio-of-lowpass (RoLP) pyramid, Daubechies wavelet four, spatially-invariant discrete wavelet transform (SIDWT), and Simoncelli's steerable pyramid. The detailed implementation of these algorithms can be found in references [64, 5, 65, 73, 69]. The basic fusion rule applied is averaging the low-frequency components while selecting the coefficients with larger absolute value in other frequency bands. The decomposition was carried out to level four and four orientational frequency bands were employed in the steerable pyramid implementation. The fused images were firstly evaluated by these criteria: RMSE, NLSE, PSNR, CORR, DE, MI, SSIM, P_{ref}, and FSSIM metric when compared with the reference image. The results are listed in Table 5.5 to 5.9.

The "best" fusion results are highlighted with bold fonts in the tables. The CORR, SSIM, FSSIM, and P_{ref} metric vote the steerable pyramid and have relative stable assessment for the different algorithms. Again, the standard deviation values are computed and listed in Table 5.10. The FSSIM metric has a relative large standard deviation value, compared with the rest.

5.6 Discussion

For the image "gold hill" and "Lena", the differences of the image quality are highlighted by the proposed metrics, namely FSSIM and P_{ref}. In terms of the standard deviation,

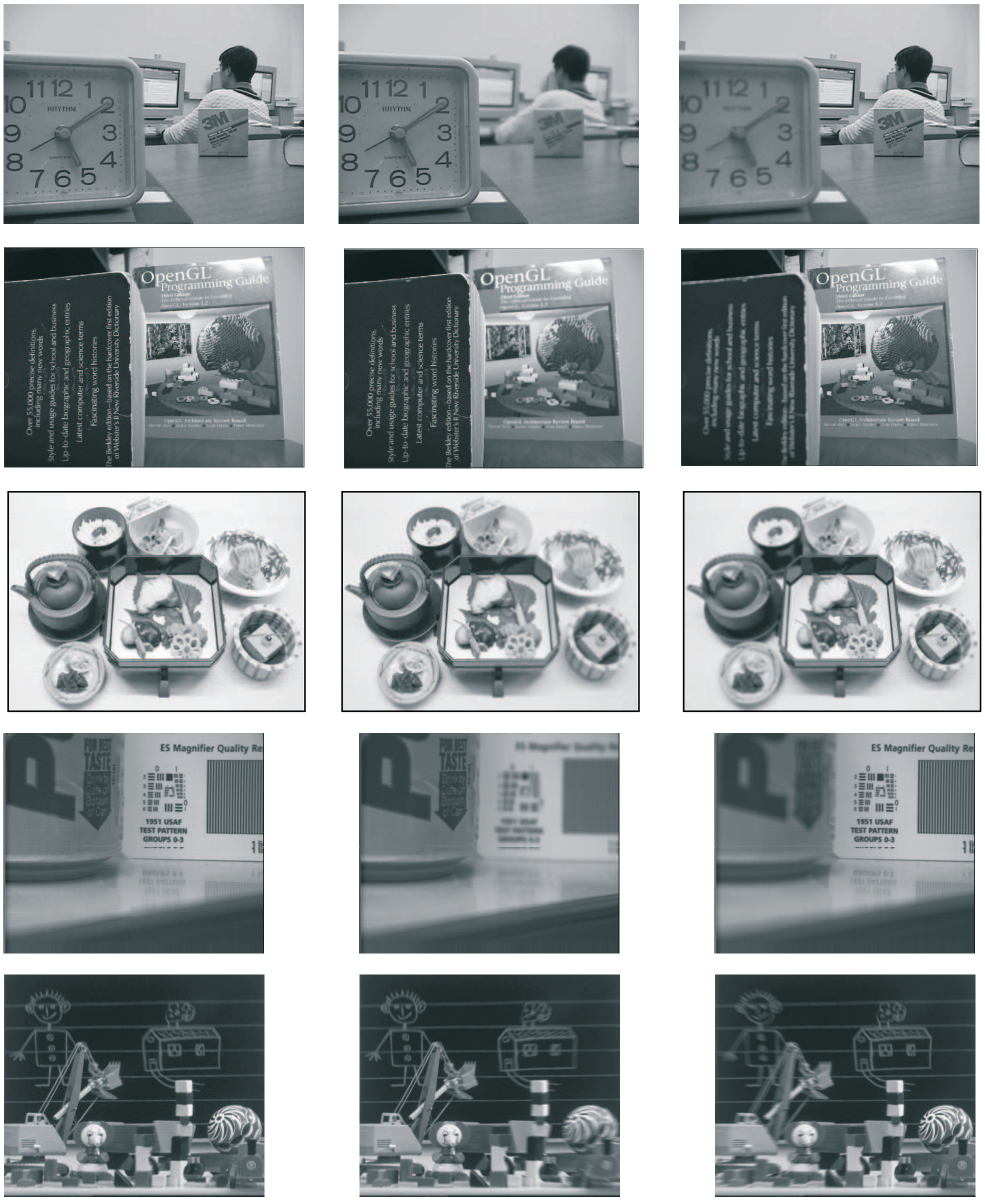


Figure 5.12: The multi-focus images used for the test. From top to bottom: laboratory, books, Japanese food, Pepsi, and object. From left to right: full-focus image, left-focus image, and right-focus image.

Table 5.5: Evaluation of the fusion result of multi-focus image “laboratory”.

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
RMSE	3.9202	7.3346	12.587	4.4012	4.386	**3.8849**
NLSE	0.0296	0.0555	0.0952	0.0333	0.0332	**0.0294**
PSNR	24.721	**28.41**	20.784	24.384	25.97	23.553
CORR	**0.9996**	0.9984	0.9956	0.9994	0.9995	**0.9996**
DE	0.0566	0.1659	0.1053	0.1497	**0.0398**	0.0870
MI	**2.4652**	2.0920	1.9992	2.3567	2.4270	2.4341
SSIM	0.9809	0.9683	0.9253	0.9762	0.9782	**0.9855**
P_{ref}	0.8297	0.8199	0.6832	0.8000	0.8212	**0.8488**
FSSIM	0.8504	0.8250	0.6706	0.8104	0.8307	**0.8559**

Table 5.6: Evaluation of the fusion result of multi-focus image “books”.

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
RMSE	5.4013	8.5444	18.3360	5.5587	5.5313	**4.5888**
NLSE	0.0511	0.0830	0.1638	0.0527	0.0524	**0.0435**
PSNR	23.2540	24.0330	22.1090	19.9820	**28.0010**	23.9520
CORR	0.9987	0.9966	0.9857	0.9986	0.9986	**0.9991**
DE	0.1473	0.0579	**0.0081**	0.3191	0.0966	0.1673
MI	2.5355	2.0977	2.2000	2.4839	2.5165	**2.7075**
SSIM	0.9556	0.9485	0.9064	0.9503	0.9538	**0.9661**
P_{ref}	0.7419	0.7344	0.6355	0.7120	0.7423	**0.7691**
FSSIM	0.7558	0.7359	0.6394	0.7138	0.7573	**0.7751**

Table 5.7: Evaluation of the fusion result of multi-focus image “Japanese food”.

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
RMSE	**9.0354**	17.0190	9.4889	9.1864	9.3845	9.1162
NLSE	**0.0517**	0.0994	0.0538	0.0527	0.0538	0.0523
PSNR	32.4400	27.9130	29.3440	32.7300	29.6740	**32.8100**
CORR	**0.9987**	0.9954	0.9986	**0.9987**	0.9986	**0.9987**
DE	0.0219	0.1964	0.0833	**0.0160**	0.0087	0.0162
MI	2.5227	1.7691	2.2717	2.5912	2.5198	**2.6265**
SSIM	0.9808	0.9370	0.9591	0.9820	0.9800	**0.9833**
P_{ref}	0.8956	0.9007	0.8554	0.9004	0.8944	**0.9076**
FSSIM	0.9035	0.8573	0.8601	0.9082	0.9011	**0.9140**

Table 5.8: Evaluation of the fusion result of multi-focus image “Pepsi”.

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
RMSE	3.4475	5.9806	10.8900	3.9439	4.6410	**3.4466**
NLSE	**0.0320**	0.0561	0.0971	0.0367	0.0431	**0.0320**
PSNR	25.4720	25.6130	25.6870	26.7070	29.1220	**30.9460**
CORR	**0.9995**	0.9984	0.9951	0.9993	0.9991	**0.9995**
DE	0.3254	0.3469	0.4090	**0.2644**	0.3651	0.3474
MI	**2.3328**	2.0024	2.0062	2.1990	2.2171	2.2983
SSIM	**0.9519**	0.9429	0.9177	0.9427	0.9464	0.9502
P_{ref}	0.4745	0.4536	0.4189	0.4670	0.4597	**0.4851**
FSSIM	0.6276	0.5907	0.5367	0.5879	0.6016	**0.6293**

Table 5.9: Evaluation of the fusion result of multi-focus image “objects”.

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
RMSE	4.7977	9.0774	10.2290	4.5204	5.3637	**4.0017**
NLSE	0.0597	0.1187	0.1217	0.0566	0.0670	**0.0501**
PSNR	**31.6830**	24.0500	29.7230	28.3720	29.0960	30.2070
CORR	0.9982	0.9931	0.9920	0.9984	0.9977	**0.9987**
DE	0.0585	0.1489	0.0886	0.1318	**0.0279**	0.0791
MI	2.0762	1.6046	1.9335	2.0273	2.0884	**2.1842**
SSIM	0.9651	0.9437	0.9382	0.9617	0.9658	**0.9720**
P_{ref}	0.7428	0.7447	0.6803	0.7301	0.7333	**0.7695**
FSSIM	0.7908	0.7646	0.7209	0.7756	0.7838	**0.8144**

Table 5.10: The standard deviation of the assessment results for the fusion of multi-focus images.

	CORR	SSIM	P_{ref}	FSSIM
Objects	0.0030	0.0135	0.0295	0.0314

the FSSIM and P_{ref} have a better performance than SSIM and other metrics, especially when the image is heavily blurred as described in Appendix B. For the reference-based assessment of a fusion result, similar conclusion can be drawn. As the fusion algorithm is concerned, the steerable pyramid appears to has the best performance for the application of multi-focus images. These multi-focus images are tested with the blind assessment metrics again in Chapter 6.

The comparison metrics reflect the image distortions from different points of view.

Each metric has its own advantages. They must be carefully selected or tailored for a specific application. Sometimes, a single metric cannot tell the "truth" and the use of multiple metrics could be a choice. The purpose of this work is not trying to propose a method that can outperform all the others in all circumstances. The advantage and feasibility of the proposed approach is illustrated and presented. For the multi-focus images, the proposed two methods give a better assessment of the fused image in terms of its standard deviation values.

Obviously, the proposed method belongs to the second category defined by Wilson. The target of the work is to find the feature difference between two images. The features considered herein may provide useful information for further analysis and processing. Actually, image quality and image comparison are two relevent but distinct topics. If there is a spatial shift of the image, for example one pixel along horizontal direction, the image is "different" from the original one, but how can we state the quality of the image has changed a lot? Actually, some currently available image quality metrics do the comparison work. Usually, the comparison needs a reference and the metric is a relative quantity while the quality metric tries to achieve an absolute measure without any reference. An universal image quality measurement is still extremely difficult to achieve. Our work presented in this chapter is limited to the comparison of image features, which are useful for further processing.

The limitation of the proposed approach is still the computational efficiency issue. Nevertheless, through using a separable approximation, the non-separable orientational filters can be decomposed into a sum of separable filters as proposed in [119]. The computational load can be reduced. Moreover, the use of field-programming technology can provide efficient solution for algorithm implementation [120]. In our implementation, we use the cross-correlation to evaluate the similarity of the feature maps. In future work, the distance measure used for Δ_g will be considered as a choice for comparison in spite of the its computational efficiency.

5.7 Conclusion

In this chapter, two new feature-based metrics for image comparison are proposed. The feasibility of this metric is investigated in the experiments with standard images and fused images. The effectiveness of the proposed metrics are demonstrated by the experimental results. The two metrics achieve a consistent assessment for the multi-focus images. Moreover, a larger deviation value indicates a better capability to identify the difference of fusion performance. For the multi-focus image experiments, Simoncelli's steerable pyramid achieved the best fusion performance in terms of the proposed metrics. These metrics can be applied to the case where image features like step edges and lines are considered. However, an individual method cannot provide a one-size-fits-all solution to all the applications. In some cases, multiple metrics need to be employed to reach an application-specified objective assessment.

The fused image can be assessed with the reference-based metrics only when the perfect reference is available, such as the multi-focus digital images. The experimental results in this chapter provide a basis for the study on blind assessment in the next chapter. In a real application, a perfect reference is not always available, for example the application in Chapter 3. In Chapter 6, the algorithms for the assessment without a reference are developed. Experiments are carried out on the fused results of both the multi-focus images and the context enhancement applications presented in Chapter 3.

Chapter 6

Feature-based Metrics for Blind Assessment

6.1 Introduction

As described in Chapter 5, a straightforward way to implement the evaluation is through the comparison with a reference image, which is assumed to be perfect. The metrics for image comparison are often employed for the quantitative assessment of image fusion. Unfortunately, the reference image is not always perfect or available practically, thus, raising the need for a quantitative and blind evaluation of the fused images.

With different purposes, the image fusion algorithms can be classified as either combination or classification fusion. In the first case, the fusion algorithm consists of combining the complementary features from multiple input images while in the second case the redundant information is mainly used for making a decision through modeling. The output of the first type of image fusion is still an image but comprised of the most salient features captured from the different sensors. The success of the post-processing or analysis relies largely on the efficiency of the specific fusion algorithm. The goal of the classification fusion is to derive a thematic map that indicates certain homogeneous characteristics of pixel regions in the image. This process needs a higher-level operation like feature- or decision-level fusion. In this case, the resulting thematic map is compared with the ground truth. The classification results are then used to generate a confusion matrix. Alternatively the classification errors for each of the classes, and for various thresholds, can also be represented by a receiver operating characteristic (ROC) curve. The perfect results can be prepared with the help of experts' experience.

A typical example for pixel-level image fusion is the fusion of multi-focus images from a digital camera [70, 75]. In such case, a cut and paste operation is applied to obtain the full-focus image that will serve as a reference for evaluating the fusion results. The evaluation metric should be optimized for image features. Pixel-by-pixel comparison does not meet the requirement, because in the original image pixels are closely related. It would be better if the quantitative evaluation can be achieved without the presence of reference image. This is the case of most practical applications. The evaluation metric should provide a measurement of how well the information of the inputs is integrated into the output.

In this chapter, three methods are proposed to identify the availability and quality of input features in the fused image. The first one is implemented by computing the local cross-correlation of the phase congruency maps between the fused and input images. A maximally selected phase congruency map is also generated for the comparison. The second one is based on the modified structural similarity measurement, where phase congruency is employed as the structural component. Similarity map with the fused image is generated for each input image. Then, the larger value at each location is retained for overall assessment. Similar to the first method, the third method considers both the phase congruency map and corresponding principal moments. The index value is obtained by averaging the cross-correlation or the similarity value in each pre-defined region. The proposed schemes achieve a no-reference evaluation of the fused image. Experiments are carried out on a group of fused images obtained by different multiresolution fusion algorithms in Chapter 3 and 5. The effectiveness of the proposed methods are demonstrated.

The rest of this chapter is organized as follows. An overview of the blind metrics used for image fusion is presented in section (6.2). In section (6.3), the concept and implementation of the feature-based evaluation are described. Experiments with the proposed approach and comparison with other existing methods can be found in section (6.4). To validate the proposed algorithms, both the reference-based approaches and no-reference methods are tested. Section (6.5) presents the discussion. In the final section (6.6), the conclusion of this chapter is drawn.

6.2 Blind Evaluation of Image Fusion

It would be better if the assessment can be accomplished without any reference image. The fused image only needs to refer to the input images to evaluate itself. Qu et al. considered mutual information (MI) and simply used the summation of the MI between the fused image (F) and inputs (A and B) to represent the difference in quality. The expression of MI-based fusion performance measure M_F^{AB} is [41]:

$$M_F^{AB} = \sum_{i,j} h_{AF}\left(i,j\right) \log_2 \frac{h_{AF}\left(i,j\right)}{h_A\left(i\right) h_F\left(j\right)} + \sum_{i,j} h_{BF}\left(i,j\right) \log_2 \frac{h_{BF}\left(i,j\right)}{h_B\left(i\right) h_F\left(j\right)} \tag{6.1}$$

where $h_{AF}\left(i,j\right)$ indicates the normalized joint grey level histogram of images A and F, $h_K\left(i,j\right)$ ($K = A, B$, and F) is the normalized marginal histogram of image A, B, or F. However, the MI metric still needs a reference value to compare with. We cannot tell in advance if a fused image with a given MI value is good or not; a reference point is a must.

Furthermore, the MI-based approach is insensitive to impulsive noise and is subject to great change in the presence of additive Gaussian noise.

Another strategy of blind assessment without reference was proposed by Xydeas et al. in [42, 43]. Their method aimed at measuring the amount of visual information transferred from the input images to the fused image. With the assumption that the edge information is closely related to the visual information, the metric is defined in terms of edge strength and orientation. The Sobel edge operator is used in the implementation to extract the strength and orientation information for each pixel. Unfortunately, the Sobel edge detection that is based on the measurement of the intensity gradient depends on image contrast and spatial magnification and hence one does not know in advance what level of edge strength corresponds to a significant feature [113, 115].

Recently, Piella and Heijmans defined a fusion quality index based on Wang and Bovik's work on universal image quality index (UIQI) (see equation (2.1)) [44]. The quality measurement is applied to local regions using a sliding window of size 8 by 8 from top left to bottom right of the image. The overall quality is given by the average $Q = \sum_{k=1}^{K} Q_b/M$, where K is the total number of image blocks. The formula can be rewritten in another form like:

$$Q_0\left(a,b\right) = \frac{1}{|W|}\sum_{w\in W} Q_0\left(a,b\left|w\right.\right) \tag{6.2}$$

where a and b are two images for comparison while w stands for the sliding window. W is the family of all windows and $|W|$ is the cardinality of W. Piella and Heijmans defined three fusion quality indexes based on the UIQI concept [44, 45]:

$$
\begin{aligned}
Q\left(a,b,f\right) &= \frac{1}{|W|}\sum_{w\in W}\left[\lambda\left(w\right)Q_0\left(a,f\left|w\right.\right)+\left(1-\lambda\left(w\right)Q_0\left(b,f\left|w\right.\right)\right)\right] && (6.3)\\
Q_w\left(a,b,f\right) &= \sum_{w\in W}c\left(w\right)\left[\lambda\left(w\right)Q_0\left(a,f\left|w\right.\right)+\left(1-\lambda\left(w\right)Q_0\left(b,f\left|w\right.\right)\right)\right] && (6.4)\\
Q_E\left(a,b,f\right) &= Q_w\left(a,b,f\right)Q_w\left(a',b',f'\right)^{\alpha} && (6.5)
\end{aligned}
$$

where f stands for the fused image of a and b, and the variance measure $\lambda\left(w\right)$ can be expressed as:

$$\lambda\left(w\right)=\frac{s\left(a\left|w\right.\right)}{s\left(a\left|w\right.\right)+s\left(b\left|w\right.\right)} \tag{6.6}$$

where there are $C\left(w\right)=\max\left(s\left(a\left|w\right.\right),s\left(b\left|w\right.\right)\right)$ and $c\left(w\right)=C\left(w\right)\big/\left(\sum_{w'\in W}C\left(w'\right)\right)$. Herein, $s\left(a\left|w\right.\right)$ and $s\left(b\left|w\right.\right)$ are the local salience of image a and b respectively. One choice is using variance of image a and b within the window w of size 8 by 8. Equation (6.3) gives a general definition of fusion quality index. The value Q_0 measures the difference between the inputs and the fused image, weighted by the variance measurement. a', b', and f' are the corresponding edge map of image a, b, and f respectively. The weighted fusion quality index actually carries out a maximum selection operation to emphasize the overall importance of each block. In equation (6.5), the edge-dependent fusion quality index tries to include the effect of edges, their contribution being controlled by the parameter α. Piella and Heijmans also implemented a weighted sum or a scaled weighted sum of the similarity measure UIQI or SSIM [44, 45]. The success of Piella's metrics depends on the performance of UIQI or SSIM.

6.3 A Strategy for the Feature-based Evaluation

This section describes the schemes for the blind assessment of fused image. The proposed feature-based strategy proceeds in two steps: first extracting image features and then measuring how those features are integrated in the fused image. The phase congruency and its principal moments are employed to provide an absolute measurement of image feature. Such measurements are incorporated into the SSIM or a local cross-correlation is performed to determine if the features from inputs are available in the fused image. An overall evaluation is obtained by averaging those local measurements.

6.3.1 Principal Moments of Phase Congruency

The principal moments of the phase congruency contain the information for the corners and edges. The magnitude of the maximum and minimum moment can be used directly to determine the edge and corner strength [121]. At each point of the image, the following are computed:

$$a = \sum \left(PC\left(\theta\right)\cos\left(\theta\right)\right)^2 \tag{6.7}$$

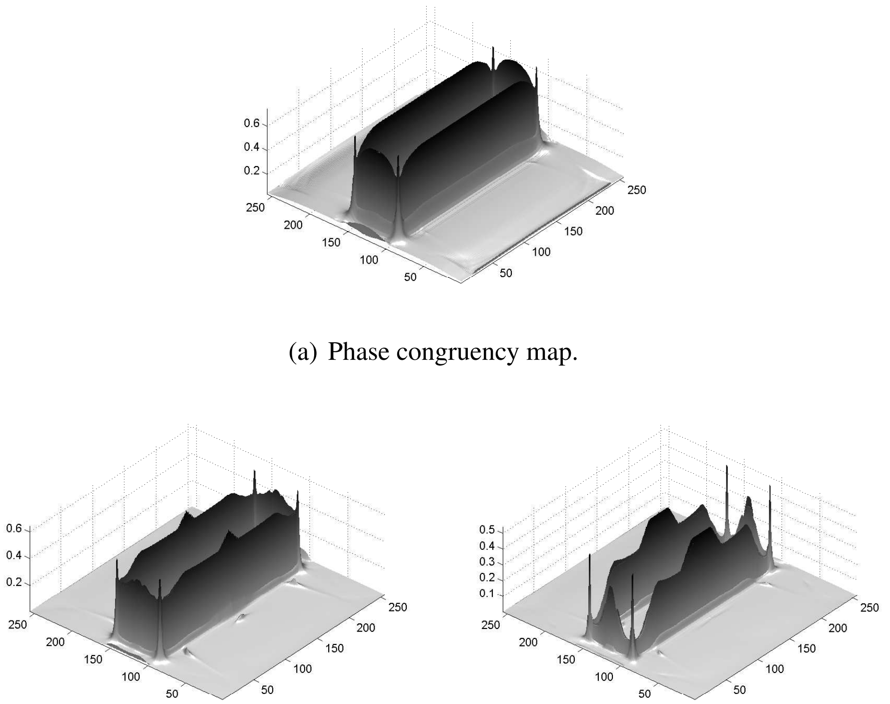
$$b = 2\sum \left(PC\left(\theta\right)\cos\left(\theta\right)\right)\cdot\left(PC(\theta)\sin(\theta)\right) \tag{6.8}$$

$$c = \sum \left(PC\left(\theta\right)\sin\left(\theta\right)\right)^2 \tag{6.9}$$

where $PC(\theta)$ is the phase congruency value along orientation θ and the sum is performed over the six directions. Therefore, the maximum (M) and minimum (m) moments are given by [121]:

$$M = \frac{1}{2}(c + a + \sqrt{b^2 + (a - c)^2}) \tag{6.10}$$

$$m = \frac{1}{2}(c + a - \sqrt{b^2 + (a - c)^2}) \tag{6.11}$$



(a) Phase congruency map.

(b) Maximum moment.

(c) Minimum moment.

Figure 6.1: The principal moments of phase congruency of the image in Figure 3.2(a).

An example is given in Figure 6.1, where the phase congruency map, maximum and minimum moment maps of the block in Figure 3.2(a) are presented. How the image features, e.g. edge and corner, are represented in these feature maps can be observed.

6.3.2 Quality Metrics for Evaluating Image Fusion

As stated previously, a blind evaluation of the fused image is preferred for practical applications, because a ground truth or a perfect reference is not always available for comparison. The pixel-level fusion is to integrate image features like edges, lines, and region boundaries into one composite image. The success of the fusion algorithm will be assessed by the measure of image features available in the fused image and those from multiple input sources. Phase congruency and its principal moments are used as the bases for the feature extraction and measurement.

The P_{blind} Metric

As proposed in Chapter 5, the first consideration is to compare the phase congruency maps of the fused image and the inputs. For the combinative fusion, the feature in the fused image may come from input images or a combination of them, as shown in Figure 6.2. The phase congruency maps of the input and fused images are firstly calculated. A third feature map M_{pc} is derived by point-by-point maximum selection of the two input maps A_{pc} and B_{pc}, i.e. retaining the larger feature points $M_{pc}\left(i,j\right) = \max\left(A_{pc}\left(i,j\right), B_{pc}\left(i,j\right)\right)$. The feature map of the fused image, F_{pc}, is then compared to A_{pc}, B_{pc}, and M_{pc} locally. At each sub-block, the cross-correlation values between these maps are computed. The evaluation index P_{blind} is the average over all the blocks. The procedure is shown in Figure 6.3.

Unlike pixels which are closely related in the original images, the points in the phase congruency map indicate the salience of image features. Therefore, the selection of feature points is not equivalent to the selection of pixels with larger intensity value in the original

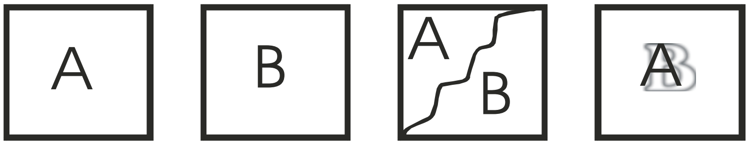


Figure 6.2: Four cases in a combinative fusion. For a small local region in the fused image, the local feature may come from the corresponding block of the input image A or B, or a combination of them.

image followed by the computation of the whole phase congruency map. Selecting larger feature points can provide a reference for comparison although such arrangement is not always optimal.

The F_{blind} Metric

Based on the FSSIM presented in Chapter 5, we propose a F_{blind} metric to assess the fused image without a reference. Mathematically, the F_{blind} can be expressed as:

$$F_{blind} = \frac{1}{K} \sum_{k=1}^{K} \left(\max \left(FSSIM_k \left(F_{pc}, A_{pc} \right), FSSIM_k \left(F_{pc}, B_{pc} \right) \right) \right) \tag{6.12}$$

where F_{pc} stands for the phase congruency map of fused image while A_{pc} and B_{pc} are the inputs'. The FSSIM between the fused image and the input images is computed respectively. The larger value means stronger feature from input images is detected in the fused image. This value is retained for final summation and average.

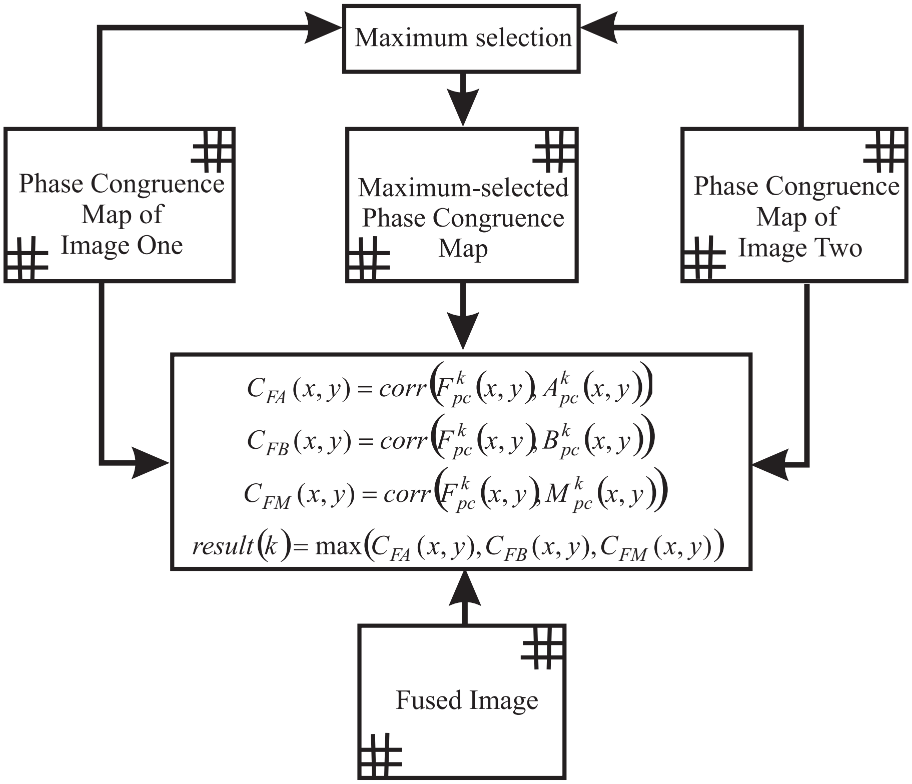

Maximum selection
Phase Congruence Map of Image One
Maximum-selected Phase Congruence Map
Phase Congruence Map of Image Two
$C_{FA}(x,y) = corr(F^k_{pc}(x,y), A^k_{pc}(x,y))$
$C_{FB}(x,y) = corr(F^k_{pc}(x,y), B^k_{pc}(x,y))$
$C_{FM}(x,y) = corr(F^k_{pc}(x,y), M^k_{pc}(x,y))$
$result(k) = \max(C_{FA}(x,y), C_{FB}(x,y), C_{FM}(x,y))$
Fused Image

Figure 6.3: The blind evaluation algorithm by using phase congruency map (P_{blind}).

The P'_{blind} Metric

The third metric is implemented by comparing both the phase congruency measurement and its principal moments. This is defined by the product of the three correlation coefficients as:

$$P'_{blind} = (P_p)^{\alpha} (P_M)^{\beta} (P_m)^{\gamma} \tag{6.13}$$

When there are two input images, we will obtain three values, e.g. P_p, P_M, and P_m respectively, which is defined as the maximum one of C_{ab}^k:

$$\begin{aligned} P_p &= max\left(C_{1f}^p, C_{2f}^p, C_{mf}^p\right) && (6.14)\\ P_M &= max\left(C_{1f}^M, C_{2f}^M, C_{mf}^M\right) && (6.15)\\ P_m &= max\left(C_{1f}^m, C_{2f}^m, C_{mf}^m\right) && (6.16) \end{aligned}$$

and there is:

$$C_{ab}^k = \frac{\sigma_{ab}^k + C_k}{\sigma_a^k \sigma_b^k + C_k} \tag{6.17}$$

Herein, C_{ab}^k stands for the correlation coefficients between two sets a and b. The symbol $\{k|p, M, m\}$ refers to the phase congruency map and its principal moments. The subfix 1, 2, m, and f correspond to the two inputs, their maximum-selected map, and the result derived from the fused image. The exponential parameters α, β, and γ can be adjusted based on the importance of the three components. In our experiments, all the three values are set to one and the small constant value C_k is selected as 0.0001.

To implement a local comparison, each pixel is compared within a 11-by-11 block in the image and only the points with a phase congruency value larger than 0.1 are used in the calculation. Assume there are total K blocks in the image, the final result is obtained by:

$$P'_{blind} = \frac{1}{K}\sum_{k=1}^{K} P'_{blind}(k) \tag{6.18}$$

6.4 Experimental Results

The experiments consist of two parts. In the first part, the proposal metrics are applied the multi-focus images used in Chapter 5, which are fused by six multiresolution algorithms. The results are listed in Table 6.1 to 6.5.

Table 6.1: Evaluation of the fusion results of multi-focus image "laboratory".

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
MI (Qu)	4.0969	3.8541	3.9855	3.9999	**4.1275**	4.1095
Xydeas	0.7585	0.7107	0.6040	0.7407	0.7581	**0.7646**
Q	0.9566	0.9490	0.9178	0.9480	**0.9627**	0.9574
Q_W	0.9325	0.8964	0.8039	0.9281	**0.9364**	0.9334
Q_E	0.8730	0.7636	0.5560	0.8641	0.8667	**0.8771**
P_{blind}	**0.8878**	0.8789	0.7638	0.8042	0.8805	0.8844
F_{blind}	0.8949	0.8658	0.7487	0.8618	0.8782	**0.8952**
P'_{blind}	0.7250	0.7110	0.4337	0.5467	0.7067	**0.7263**

Referring to the results in Chapter 5, the steerable pyramid is favored by most of the reference-based metrics. In the examples of image fusion without any references, only a blind evaluation can be carried out. It is not surprised to see that the fusion algorithms exhibit different performance on different images. The steerable pyramid is selected as the best by the proposed metrics as highlighted in Table 6.1 to 6.5. The only exception is the P_{blind} metric, which prefers the Laplacian pyramid for image "laboratory" and "books". Both the reference-based and blind metrics indicate that the steerable pyramid achieves a fused image with higher quality. The proposed metrics appear to be consistent with the reference-based assessment.

Table 6.2: Evaluation of the fusion results of multi-focus image "books".

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
MI (Qu)	4.4115	3.9617	4.6247	4.3229	4.4521	**4.6372**
Xydeas	0.7348	0.6718	0.6256	0.7154	**0.7427**	0.7380
Q	0.9474	0.9400	0.9253	0.9365	**0.9569**	0.9489
Q_W	0.9332	0.8942	0.7992	0.9283	**0.9363**	0.9347
Q_E	0.8823	0.7686	0.5785	0.8682	0.8778	**0.8859**
P_{blind}	0.9001	0.8927	0.8214	0.8677	**0.9025**	0.8969
F_{blind}	0.8670	0.8417	0.7517	0.8212	0.8697	**0.8741**
P'_{blind}	0.7572	0.7461	0.5536	0.6893	0.7594	**0.7634**

Table 6.3: Evaluation of the fusion results of multi-focus image "Japanese food".

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
MI (Qu)	4.4781	3.3620	**4.5679**	4.4824	4.4813	4.5301
Xydeas	0.8910	0.8362	0.8873	0.8889	**0.8984**	0.8936
Q	0.9803	0.9506	0.9771	0.9794	**0.9829**	0.9807
Q_W	0.9706	0.9389	0.9655	0.9706	**0.9744**	0.9710
Q_E	0.9193	0.8419	0.8963	0.9188	**0.9271**	0.9203
P_{blind}	0.9734	0.9644	0.9605	0.9725	0.9731	**0.9754**
F_{blind}	0.9683	0.9175	0.9402	0.9629	0.9671	**0.9707**
P'_{blind}	0.9473	0.9242	0.8950	0.9470	0.9470	**0.9560**

In the second part of the experiment, the fused images from Chapter 3 are evaluated. The six multiresolution algorithms in Chapter 5 are applied again. In Figure 6.4 to 6.12,

Table 6.4: Evaluation of the fusion results of multi-focus image "Pepsi".

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
MI (Qu)	**4.3611**	4.0215	4.3146	4.1514	4.2510	4.3199
Xydeas	0.8180	0.7880	0.6639	0.7997	0.8143	**0.8275**
Q	0.9667	0.9627	0.9342	0.9607	**0.9711**	0.9664
Q_W	**0.9604**	0.9250	0.7910	0.9569	0.9602	0.9603
Q_E	0.9240	0.8133	0.4854	0.9153	0.9186	**0.9252**
P_{blind}	0.9185	0.9010	0.7993	0.8923	0.9043	**0.9194**
F_{blind}	0.8805	0.8543	0.7869	0.8440	0.8811	**0.9053**
P'_{blind}	0.8166	0.7853	0.5412	0.7566	0.7853	**0.8241**

Table 6.5: Evaluation of the fusion results of multi-focus image "objects".

Assessment metric	Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
MI (Qu)	3.9174	3.1582	**4.2260**	3.7720	4.0264	4.1063
Xydeas	0.7833	0.7111	0.7593	0.7524	**0.7938**	0.7853
Q	0.9629	0.9450	0.9559	0.9542	**0.9693**	0.9637
Q_W	0.9436	0.9118	0.9055	0.9389	**0.9497**	0.9457
Q_E	0.8671	0.7694	0.7311	0.8553	0.8710	**0.8745**
P_{blind}	0.9379	0.9310	0.9063	0.9147	0.9334	**0.9390**
F_{blind}	0.9110	0.8631	0.8559	0.8745	0.9012	**0.9163**
P'_{blind}	0.7973	0.7835	0.7004	0.7385	0.7946	**0.8115**

the results are arranged as shown in Table 6.6. Again, the numerical results are listed in Table 6.7 and 6.8. In this case, the assessment metrics favor the SIDWT algorithm.

Table 6.6: The arrangement of images in Figure 6.4 to 6.12.

Laplacian pyramid	Gradient pyramid
Ratio-of-lowpass pyramid	Daubechies wavelet four
SIDWT	Steerable pyramid

Although the three proposed metrics come to the same conclusion, the P'_{blind} has a relatively large standard deviation values in the two tests. Therefore, it is more appropriate to the fusion assessment application. The proposed blind metrics are all based on the phase congruency measurement and the basic principle is the same. The difference is that the F_{blind} only refers to the two input images and the maximum-selected phase congruency map from the inputs is not considered. In terms of the results from multi-focus imaging and night vision application, no big difference has been observed.

6.5 Discussion

The image fusion is application dependent. In other words, the process depends on the type of images or their formats. Because the images acquired by heterogeneous sensors possess a different intensity map, there is no “one size fits all” solution for the evaluation process. Therefore, one fusion algorithm may not necessarily achieve the same performance on distinct images in terms of certain evaluation metric. One purpose of this study is to identify the feasibility and validity of the evaluation algorithms, i.e. how these metrics work for

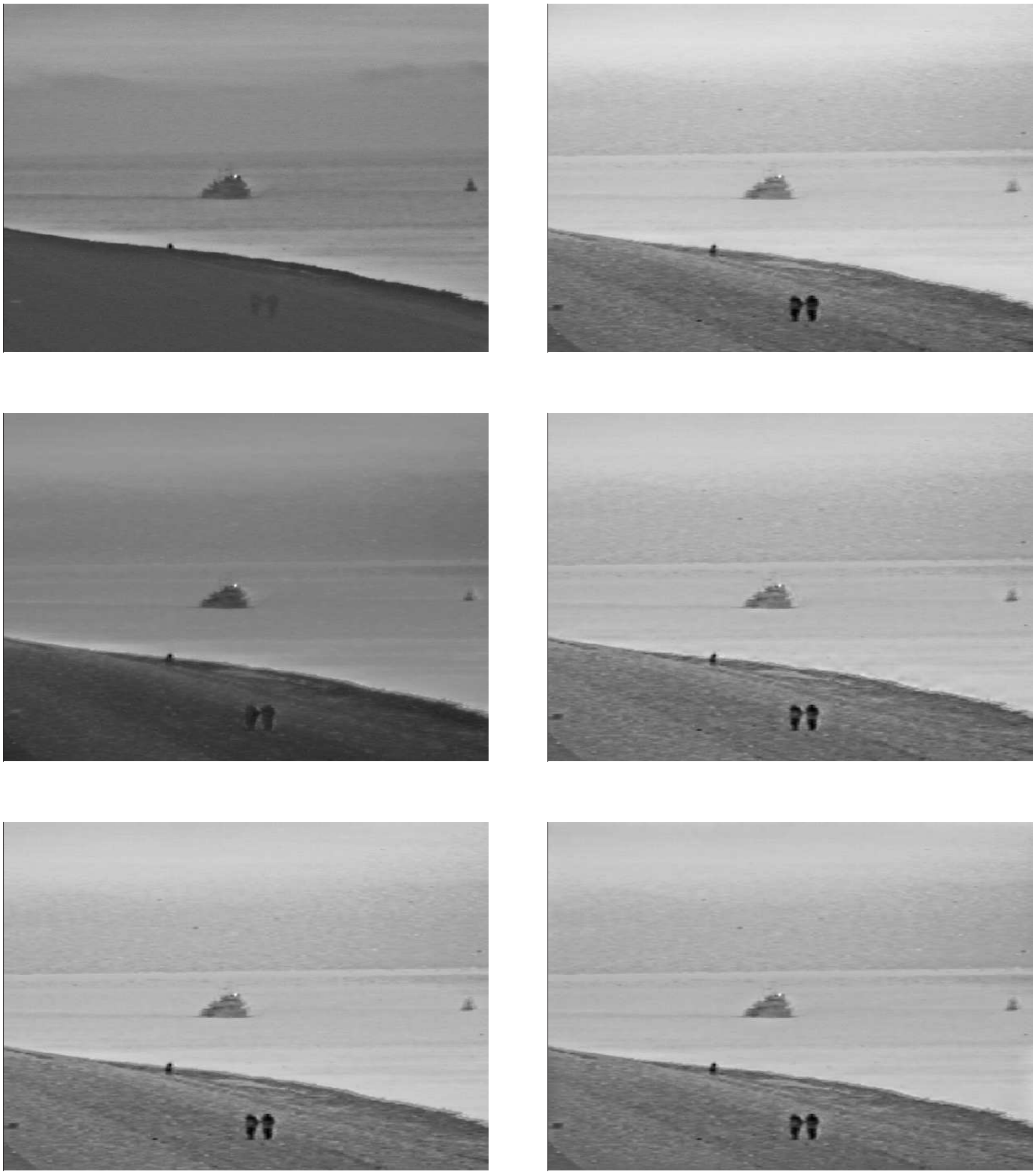


Figure 6.4: Fusion results of image “B7118”.

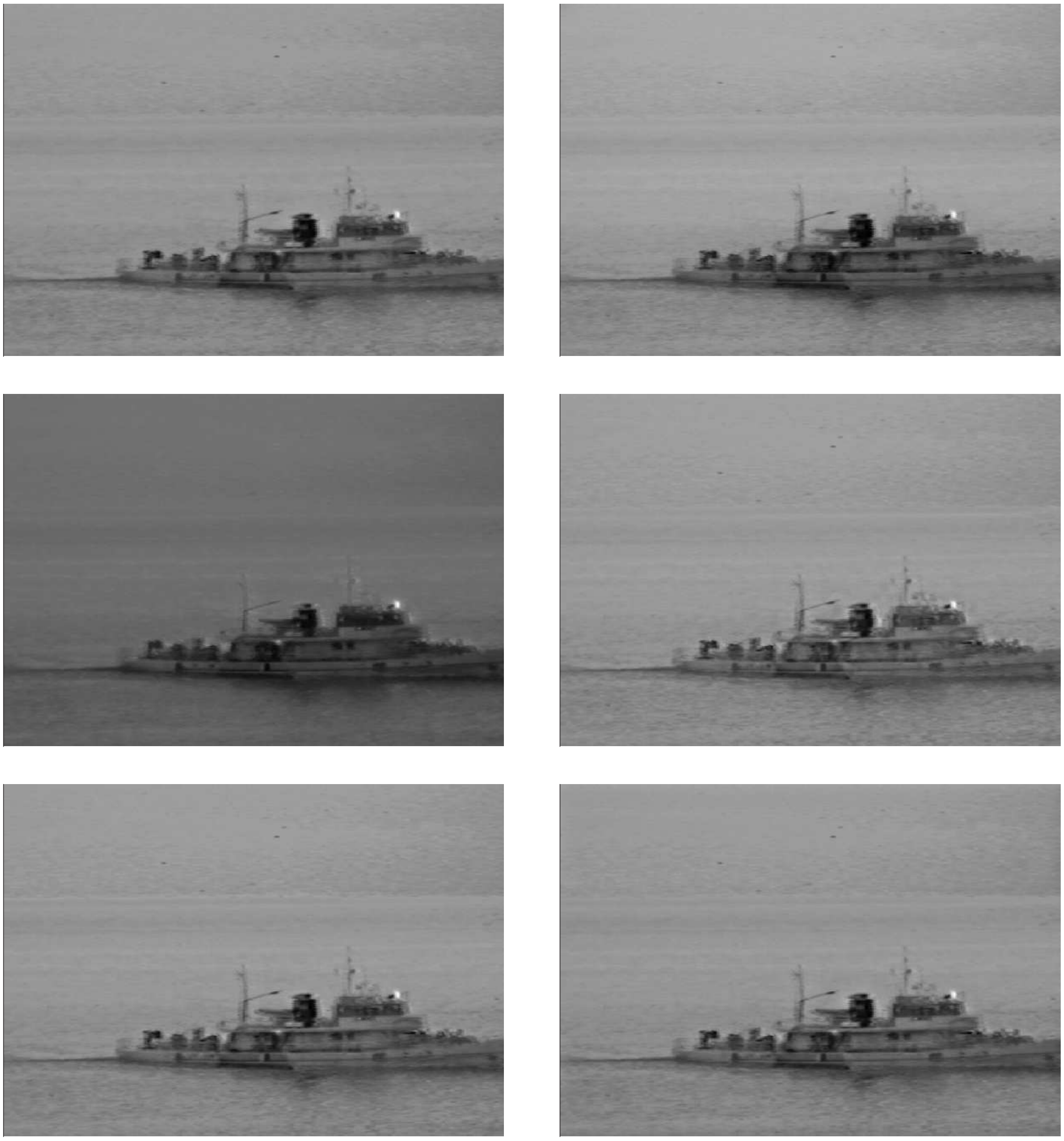


Figure 6.5: Fusion results of image “B7436”.

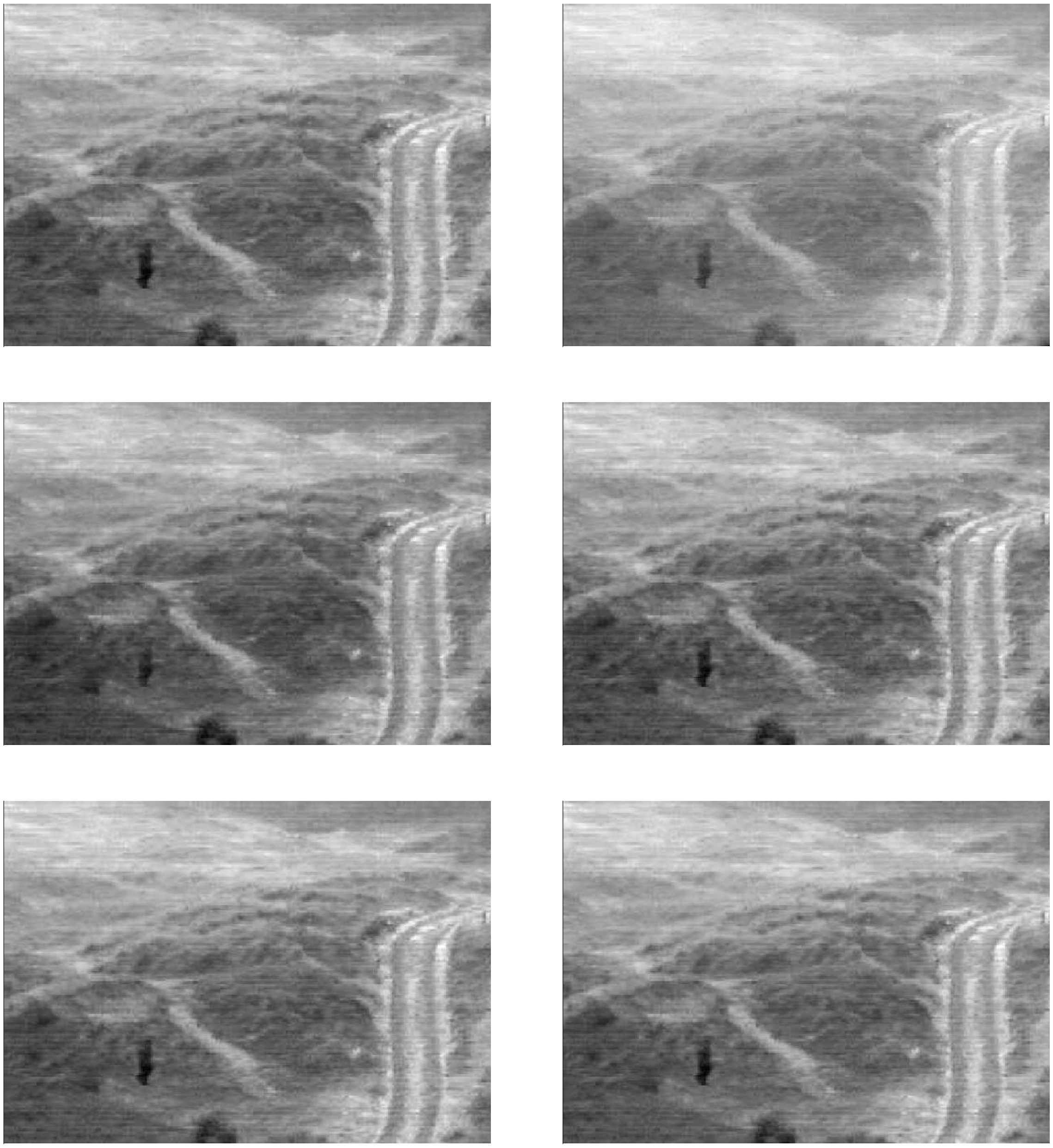


Figure 6.6: Fusion results of image “Dune”.

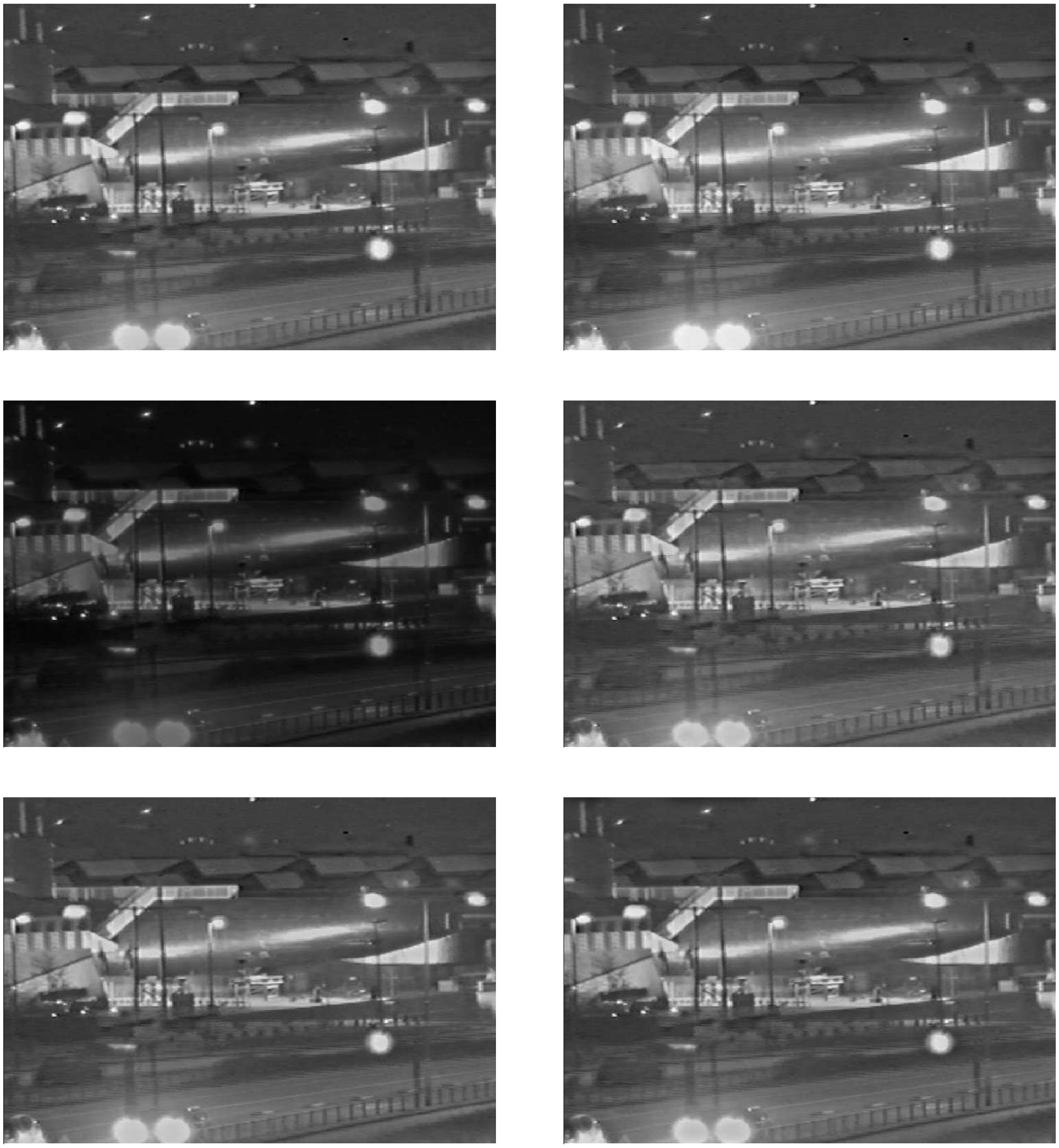


Figure 6.7: Fusion results of image “e518a”.

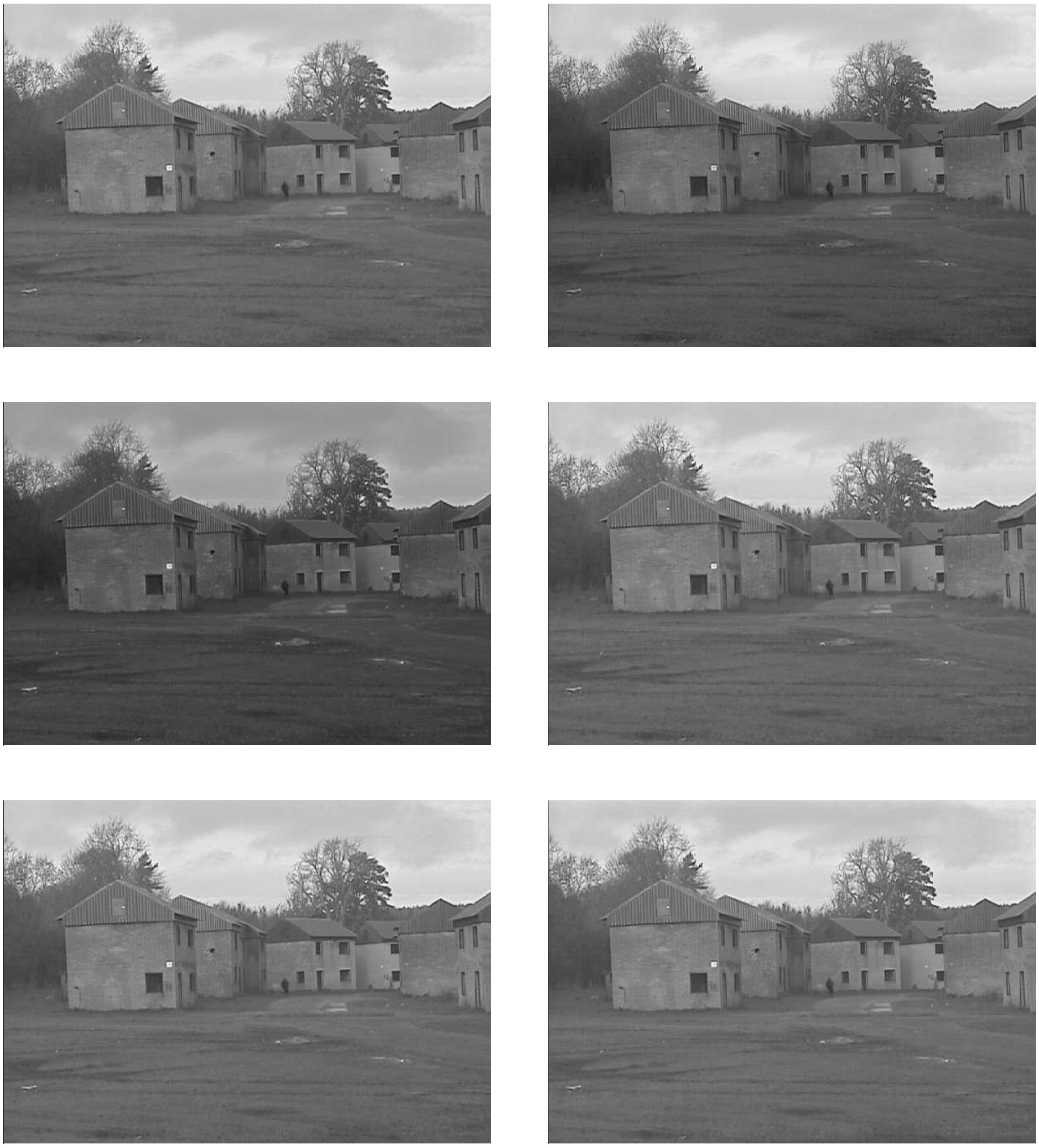


Figure 6.8: Fusion results of image “Octec02”.

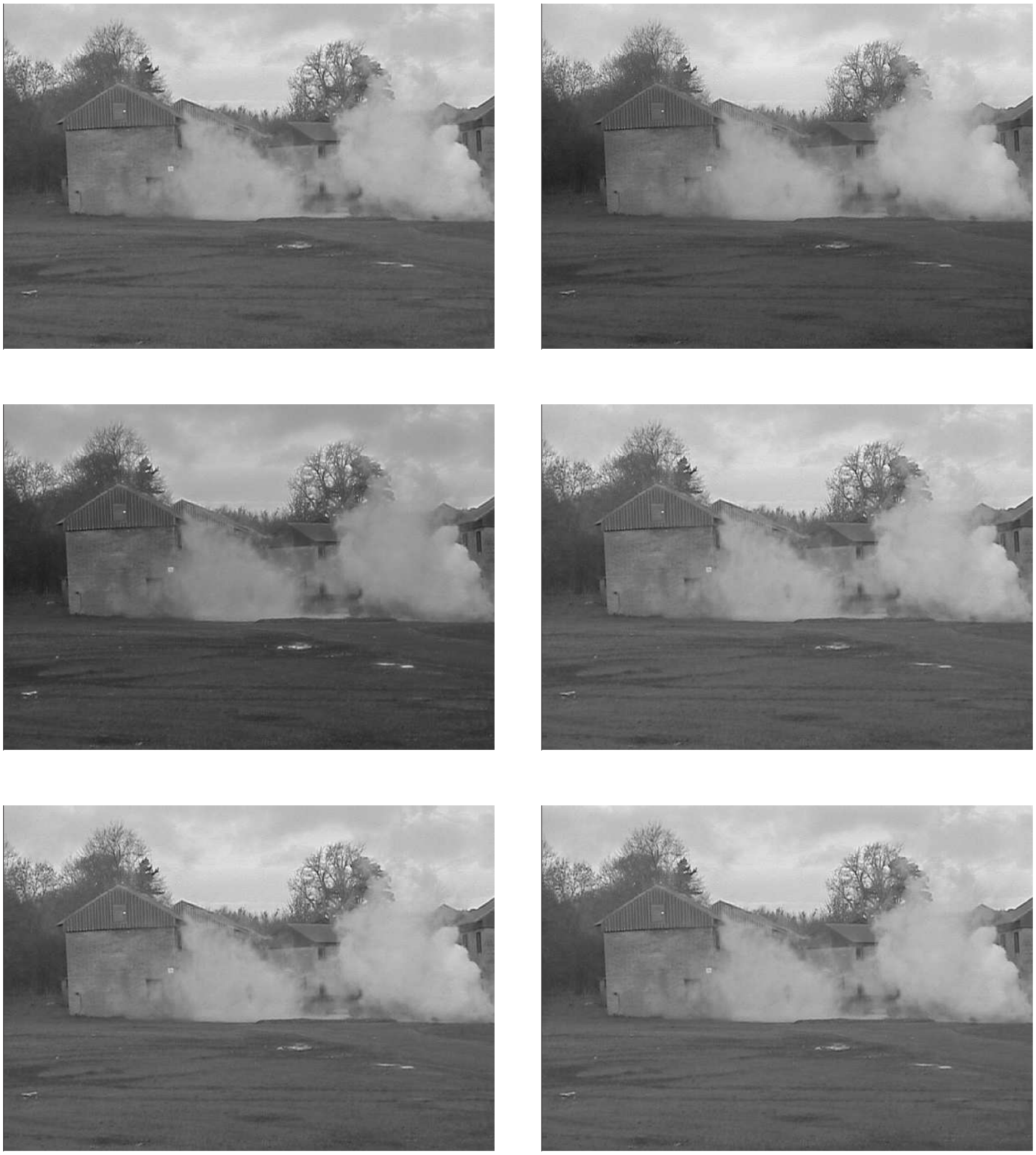


Figure 6.9: Fusion results of image “Octec21”.

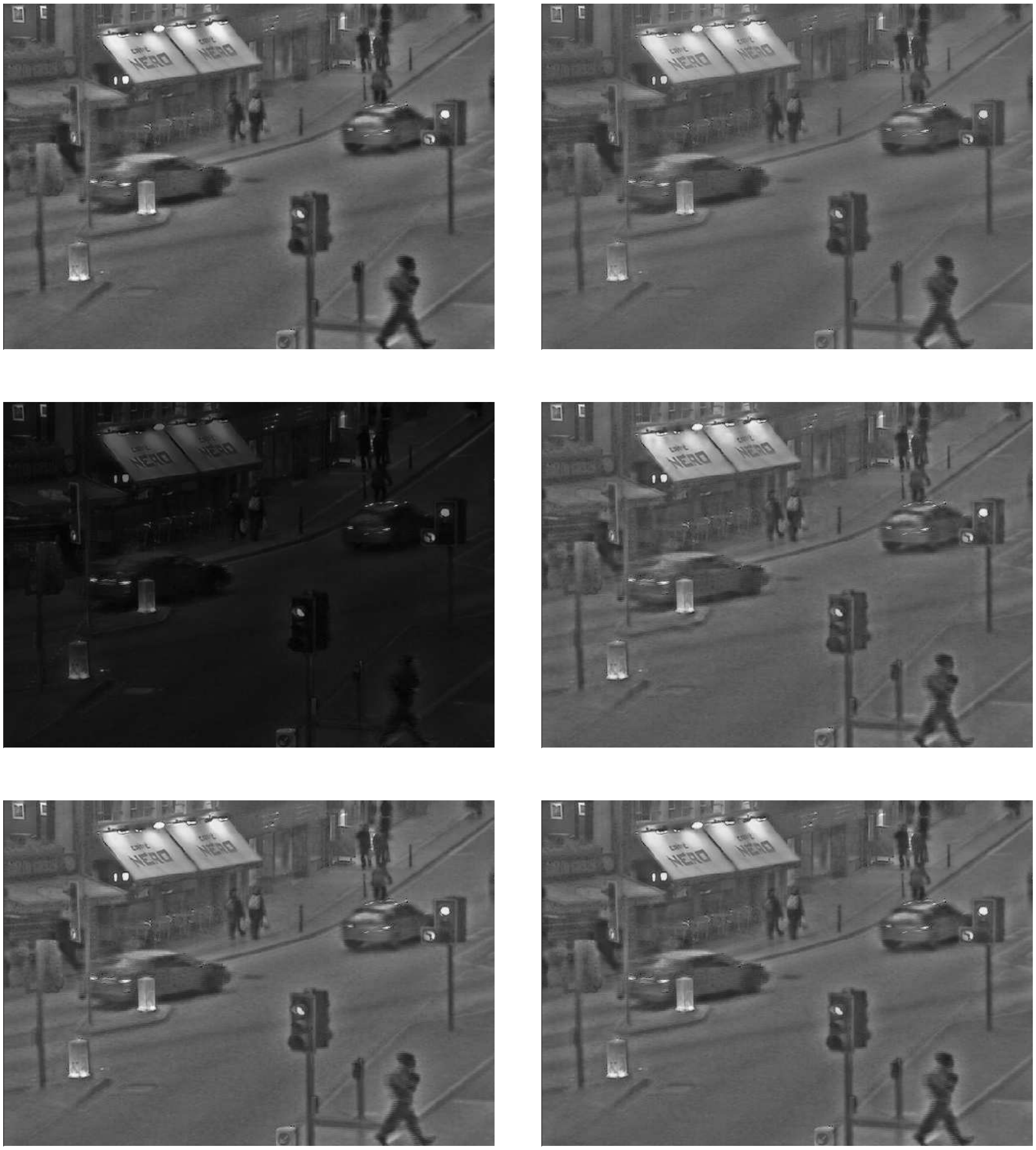


Figure 6.10: Fusion results of image “Quad”.

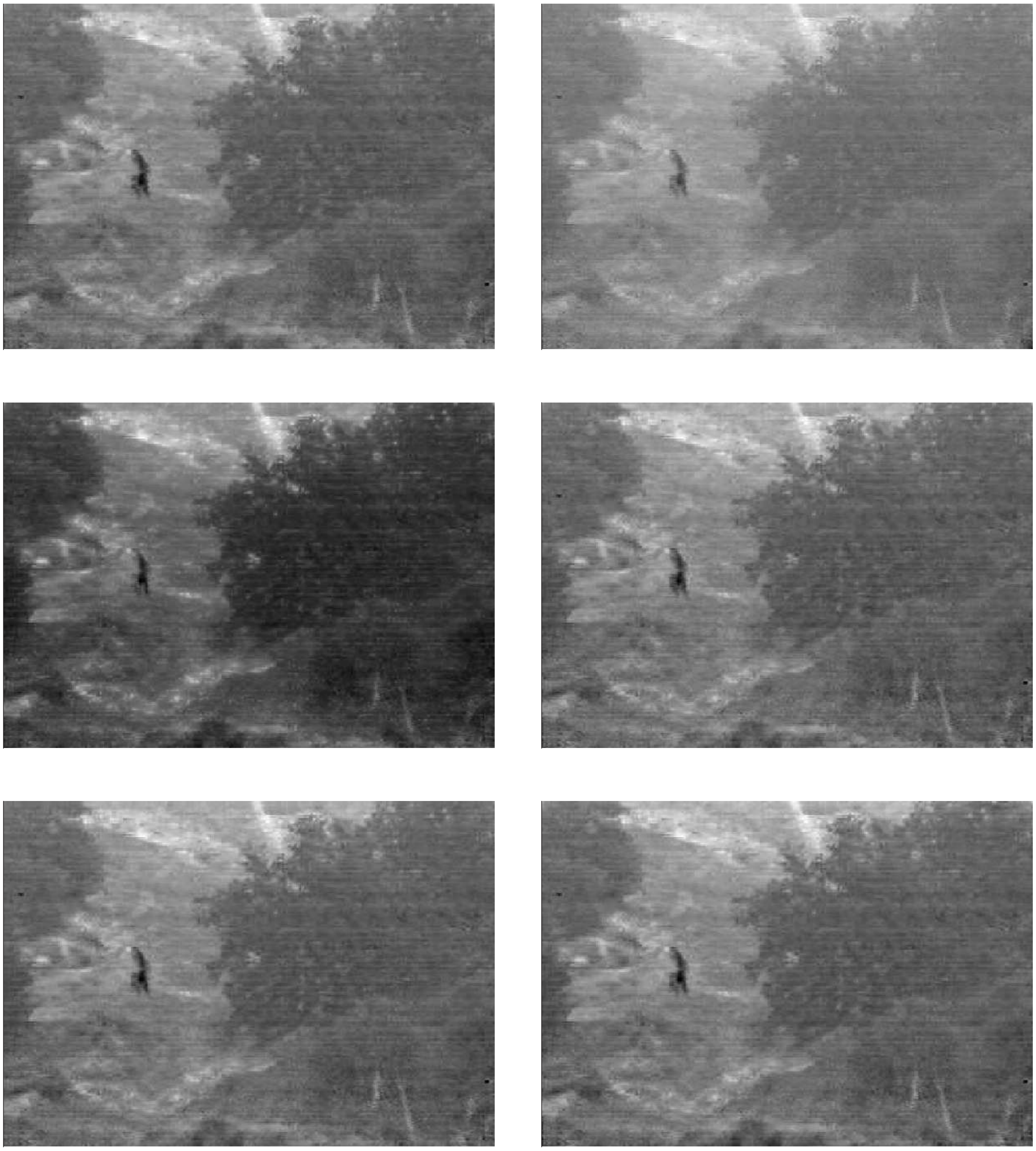


Figure 6.11: Fusion results of image “Tree 4906”.

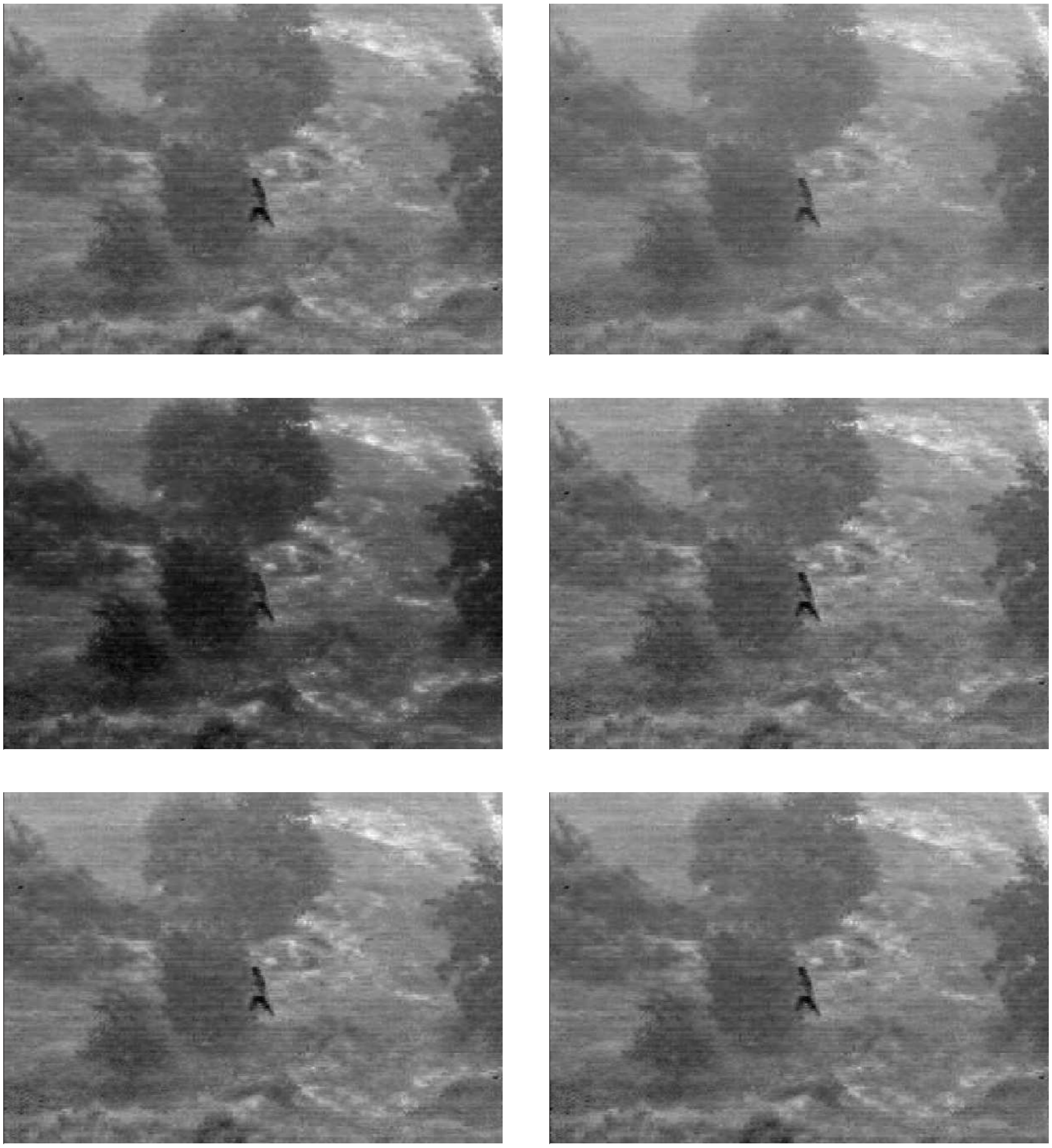


Figure 6.12: Fusion results of image “Tree 4917”.

Table 6.7: Evaluation of the fusion results of night vision images with MI, Xydeas' method, Q metrics.

Assessment metric		Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
B7118	MI (Qu)	2.7212	**3.0929**	2.7681	2.701	2.8233	2.7074
	Xydeas	0.3023	0.7384	0.5369	0.7602	**0.8004**	0.7802
	Q	0.3104	0.8504	0.7640	0.8577	**0.8800**	0.8741
B7436	MI (Qu)	2.1542	**2.5186**	2.1515	2.1537	2.2498	2.1686
	Xydeas	0.7571	0.7202	0.6114	0.7165	**0.7659**	0.7511
	Q	0.8924	0.8776	0.8145	0.8806	**0.8998**	0.8967
Dune	MI (Qu)	3.6780	3.5151	3.7214	3.6329	**3.6873**	3.6560
	Xydeas	0.8976	0.8725	0.8962	0.8820	0.8980	**0.9001**
	Q	0.9841	0.9760	0.9827	0.9822	**0.9851**	0.9849
e518a	MI (Qu)	2.7032	**2.8678**	2.4046	2.5714	2.7429	2.5685
	Xydeas	**0.7879**	0.7667	0.6802	0.7490	0.7851	0.7761
	Q	0.9224	0.9030	0.8637	0.9202	**0.9246**	0.9238
Octec02	MI (Qu)	3.8177	3.7346	3.5317	3.8158	**3.8565**	3.6789
	Xydeas	0.9088	0.8745	0.8705	0.8987	**0.9094**	0.9048
	Q	0.9576	0.9423	0.9321	0.9569	**0.9579**	0.9577
Octec21	MI (Qu)	4.2538	4.2256	4.0208	4.2444	**4.2920**	4.1500
	Xydeas	0.8880	0.8621	0.8496	0.8747	**0.8890**	0.8841
	Q	0.9560	0.9436	0.9315	0.9552	**0.9564**	0.9561
Quad	MI (Qu)	2.0516	**2.1775**	1.5454	1.7700	1.9839	1.7756
	Xydeas	**0.8078**	0.7457	0.5184	0.7368	0.7901	0.7840
	Q	0.8395	0.8451	0.7358	0.8485	**0.8551**	0.8505
Tree 4096	MI (Qu)	2.1060	**2.2028**	2.1213	2.0955	2.1420	2.0524
	Xydeas	0.8019	0.7955	0.7787	0.7772	**0.8035**	0.7982
	Q	0.9555	0.9527	0.9439	0.9519	**0.9583**	0.9571
Tree 4917	MI (Qu)	2.5658	**2.6820**	2.5495	2.5372	2.5812	2.5096
	Xydeas	0.8180	0.8107	0.7965	0.7951	**0.8201**	0.8180
	Q	0.9617	0.9577	0.9471	0.9586	**0.9641**	0.9633

Table 6.8: Evaluation of the fusion results of night vision images with proposed metrics (P_{blind}, F_{blind}, and P'_{blind}).

Assessment metric		Lapacian Pyramid	Gradient pyramid	Ratio-of-lowpass pyramid	Daubechies wavelet four	SIDWT (Haar)	Steerable pyramid
B7118	P_{blind}	0.7037	0.8442	0.6425	0.8044	**0.8583**	0.8411
	F_{blind}	0.7756	0.8314	0.6420	0.7905	**0.8479**	0.8299
	P'_{blind}	0.3790	0.6240	0.2511	0.5255	**0.6498**	0.6128
B7436	P_{blind}	0.8383	0.8265	0.6405	0.7836	**0.8425**	0.8263
	F_{blind}	0.8382	0.8286	0.6743	0.7859	**0.8435**	0.8287
	P'_{blind}	0.5900	0.5734	0.2525	0.4754	**0.5982**	0.5657
Dune	P_{blind}	0.9392	0.9272	0.9228	0.9223	**0.9402**	0.9379
	F_{blind}	0.9477	0.9277	0.9308	0.9285	**0.9483**	0.9470
	P'_{blind}	0.8247	0.7991	0.7792	0.7792	**0.8286**	0.8236
e518a	P_{blind}	0.9021	0.8965	0.8006	0.8678	**0.9034**	0.8946
	F_{blind}	0.8511	0.8324	0.7133	0.7960	**0.8507**	0.8340
	P'_{blind}	0.7561	0.7470	0.5333	0.6661	**0.7573**	0.7361
Octec02	P_{blind}	0.9600	0.9477	0.9217	0.9506	**0.9617**	0.9578
	F_{blind}	0.9018	0.8722	0.7882	0.8750	**0.9056**	0.8973
	P'_{blind}	0.9018	0.8722	0.7882	0.8750	**0.9056**	0.8973
Octec21	P_{blind}	0.9011	0.9340	0.9011	0.9337	**0.9496**	0.9440
	F_{blind}	0.9273	0.9061	0.8740	0.9083	**0.9286**	0.9232
	P'_{blind}	0.8642	0.8356	0.7378	0.8299	**0.8711**	0.8592
Quad	P_{blind}	0.8606	0.8536	0.7561	0.8124	**0.8556**	0.8423
	F_{blind}	0.7960	0.8076	0.5383	0.7288	**0.7989**	0.7697
	P'_{blind}	0.6808	0.6683	0.4545	0.5540	**0.6622**	0.6284
Tree 4096	P_{blind}	0.8475	0.8387	0.7984	0.8111	**0.8557**	0.8448
	F_{blind}	0.8636	0.8474	0.8090	0.8273	**0.8723**	0.8613
	P'_{blind}	0.6147	0.6061	0.5122	0.5419	**0.6315**	0.6097
Tree 4917	P_{blind}	0.8798	0.8719	0.8402	0.8504	**0.8859**	0.8812
	F_{blind}	0.8939	0.8770	0.8452	0.8609	**0.9003**	0.8967
	P'_{blind}	0.6855	0.6722	0.5891	0.6167	**0.6996**	0.6894

different images, rather than rank these methods in a general sense. For a particular application there should be an optimal solution to the pixel-level fusion process. However, a benchmark must be setup for such comparison in a predefined situation. To our knowledge, most work on pixel-level image fusion employs multiples metrics for assessing the fused results rather than relies on one metric only.

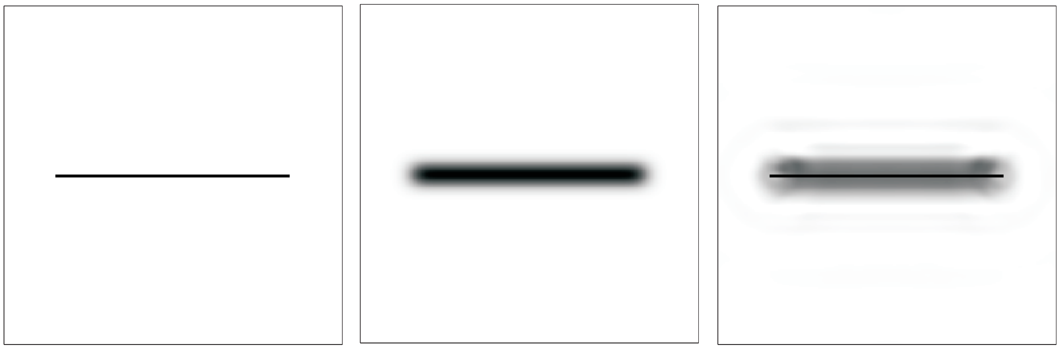
The fusion quality metrics provide a scale to assess the result and guide the choice or the structure of fusion algorithms. This could be generalized in a two-step procedure: 1) make clear what are expected in the fusion result and select one or multiple evaluation metrics for this purpose; 2) test fusion algorithms with the specific evaluation metrics and choose the appropriate one. Besides, the requirements of post-processing provide a direct test on the quality of the fused image even though this procedure may not always provide a quantitative evaluation for the fused results.

The fusion quality metrics we proposed in this chapter, i.e. P_{blind}, F_{blind} and P'_{blind}, are feature-oriented approaches. These metrics can successfully identify the image quality based on the feature measurement with the phase congruency method. The gradient-based algorithm for feature detection is inadequate for edges composed of combinations of steps, peaks, and roofs [113, 115]. The invariant qualities in images are very important to evaluate wide classes of images, which provide a very dynamic and unstructured environment for the algorithms applied [113, 114]. Phase congruency allows edges, lines, and other features to be detected reliably [113, 114, 115] and the match between the fused and input images can be detected by using the local correlation of the phase congruency as the proposed metrics. In other words, when the target of the fusion is to combine the features like step edges, lines, and Mach bands from multiple input images, the F_{blind} and P_{blind} metrics provide an effective way to assess the feasibility of the potential algorithms. In the implementation of P_{blind} metric, the cross-correlation is employed to measure the similarity of image features. Other similarity measure presented in [122] will also be considered in the future work.

It should be mentioned that the proposed approach is subject to the presence of noise, which is introduced during the image acquisition. The evaluation metrics cannot identify and remove the noise automatically. In the MRA-based pixel-level fusion, the fusion rule is to keep the coefficients with a larger absolute value, which corresponds to the image feature like lines, edges, and boundaries. The input images come with the "features" (noise) that may not be part of the perfect result; however, the coefficient selection process may eventually retains such "feature" in the fused result. We will still get a confirm from our metric that that "feature" is available in the fused image. In that sense, the assessment metric is a tool to evaluate how the information (feature) is transferred to the fused result. This does not assure a perfect result unless the features are totally complementary. That is the limitation of all approaches that are based on feature measurement in the fused image. The only solution to this problem is the optimization of fusion process rather than the assessment metric. An example is given in Figure 6.13. Image in Figure 6.13(a) has a two-pixel-wide line across the center with a gray scale value of five. The second image is the blurred version of the first image by applying the averaging operation. The maximum gray scale value is around 1.2. The two images can represent a segment from a bigger picture captured by two image modalities. The MRA-based fusion generates a result shown in Figure 6.13(c) [1].

If the evaluation metric is to assess whether the features from the two images are fused in the final result. The conclusion could be that the image in Figure 6.13(a) is "better" than the one in Figure 6.13(b). A good implementation of fusion will not introduce any noises or artifacts to the result, but the algorithm must be intelligent enough to identify what should be retained. More sophisticated fusion algorithms considers not only an isolated pixel but also its surroundings and correspondences in other frequency bands. However,

[1] The grayscale is adjusted ($[0, 0.2]$)to show the details of the result.



(a) Salient feature.

(b) Weak feature.

(c) The fused result.

Figure 6.13: The example of fusing a strong and a weak feature.

no suppression operation has been taken into account when the situation in Figure 6.13 happens. Therefore, the objective evaluation metric cannot be implemented through finding a perfect result as the reference. This is the ultimate goal of fusion. The assessment of the fusion is carried out by evaluating how the features are transferred from the inputs to the fused result.

6.6 Conclusion

In this chapter, feature-based metrics for blind assessment of image fusion performance are presented. The metrics are based on a modified SSIM scheme and the local cross-correlations between the feature maps of the fused and the input images. The image features are represented by a dimensionless quantity ranging from zero to one namely phase congruency, which is invariant to the changes in image illumination and contrast. These metrics provide an objective quality measure of the fused image in the absence of a reference image. In this chapter, the proposed blind metrics were first tested by the multi-focus images, where “perfect” references were available for comparison. The results indicate that among the three blind metrics, F_{blind} and P'_{blind} have a consistent assessment for different images. In the second application, i.e. context enhancement, the SIDWT algorithm was regarded as the best by all the three metrics.

Although image fusion is an application-dependent process, the metrics proposed in this study can be applied to the case where the features like step edges and lines are to be integrated from multiple images. The effectiveness of the approach can also been seen from the comparison with other solutions. Another interesting aspect of this work is how the metric can be utilized to optimize the fusion process at the algorithm development stage; this remains a topic for future investigations. There is no uniform standard for all image fusion applications. The evaluation metric must be chosen carefully based on the sensor type, image format, and the requirements of the particular application. In a complicated case, multiple metrics should be considered. As the fusion performance is concerned, one should mention the image modalities as well as the assessment metric being used.

Chapter 7

Conclusions

THIS thesis explores the processing of multi-modal images for surveillance applications. Advanced surveillance systems have already benefited from multiple imaging modalities, which provide information across the electromagnetic spectrum. The study presented in this thesis focuses on the registration of multi-modal video sequences (images), enhancement and characterization through fusing multi-sensor images, and the objective evaluation of combinative image fusion at pixel level. The registration, fusion, and evaluation constitute the essentials of a multi-sensor imaging system. Chapter 2 proposed a solution for registering infrared and electro-optical video sequences. Such operation assures the accuracy for pixel-based operation. The requirements of a specific application should be referred when the fusion operation is considered. Two typical applications are considered in this thesis. One is the context enhancement (Chapter 3) and the other is the concealed weapon detection (Chapter 4). The multiresolution analysis based fusion algorithms are developed and validated through experiments. The objective assessment of the performance of a fusion algorithm is addressed from two perspectives, i.e. with and without a reference. The proposed metrics will guide the choice or even optimization of a fusion algorithm for a specific application. The contributions of the thesis can be summarized as follows:

- The registration of infrared and electro-optic video sequence is carried out by solving the least square solution from matching the trajectory of head top points in consecutive frames. Matching single point is much easier than matching multiple feature points between frames, although multiple matches can be implemented. The proposed method uses the silhouette of frame difference detected by image structural similarity measurement and does not rely on the success of any foreground detection algorithms. A predefined threshold value can be applied in spite of the image modality. The initial registration parameters are refined by applying the maximum mutual information method to search the optimal result.

- The use of infrared imaging helps improve the awareness of environment under unadequate illumination. Current available techniques suggest adaptive enhancement or direct pixel-level fusion of infrared and visual images. However, the enhancement algorithm may not be truly adaptive and the derived result is not optimal for human perception, because the fusion algorithm simply integrates the features available in two spectrum ranges. In this thesis, a modified fusion scheme is proposed to process the visual image for context enhancement. The visual image is first enhanced by the infrared counterpart. The fusion of the enhanced image and the visual image is to preserve the features presented in the visible band of the spectrum.

- Chapter 4 clarifies the requirements for the application of concealed weapon detection. A new strategy is proposed to implement the fusion of multi-sensor images for concealed weapon detection. A scheme to synthesize a composite image with the information of both the personal identification (facial pattern) and concealed weapon

[0] Some parameters need to be tune manually.

is presented with the experimental results obtained from processing images acquired by CCD and infrared cameras. This study also provides a solution for privacy protection, where the concealed weapon is highlighted without exposure of human body.

- This thesis investigates two types of metrics for the assessment of combinative pixel-level image fusion algorithms, e.g. reference-based metric and blind assessment metric. The objective evaluation is based on a type of image feature measurement, namely, phase congruency. The first type is a reference-based metric, i.e. a "perfect" reference is available for comparison. The second type is a blind metric, which does not need any prior knowledge about the final result. The efficiency of the fusion algorithm is assessed through measuring the availability of features of inputs in the fused result. The proposed methods implement the quantitative evaluation.

This thesis constructs a basis for future research. Registration is the first step to process multi-modal images. How the accuracy of the registration affects the further analysis is not addressed in this thesis. The topics on how the fusion of multiple imaging modalities can facilitate the detecting, tracking, and characterizing issues in surveillance applications will be of great interest for further study.

The context enhancement through fusion is implemented with a simple exponential function. The infrared image can be further processed, for example, segmented as foreground and background, before being used to enhance the visual image. Is there any other function optimal for the enhancement? This is worth a further investigation.

For the application of concealed weapon detection, there still lacks two types of studies. The first is the "true" fusion for improved weapon detection, i.e. fusing the information from multiple detecting sensors. The second is the performance study in terms of probability of detection (POD). There are a number of factors affecting the performance of the weapon detection system, such as distance, thickness of the coat, temperature, etc. The

POD should be investigated for a specific technique or the fusion result derived from multiple techniques. If one technique has a hundred percent detection rate, it does not make any sense to fuse the result from another technique.

The new metrics for assessing the performance of image fusion algorithms have their advantages and do not exclude other metrics, because different metrics provide different measurements and reflect different characteristics. It may be necessary to refer to multiple metrics when a specific application is considered. The choice of these metrics depends on the requirements of the application.

However, this study is still limited by the availability of multi-sensor images. Collection of representative images and video sequences is difficult. Further study can be carried out when more representative data is available.

Appendix A

The Implementation of Phase Congruency Algorithm

This section is to give the information on how the phase congruency algorithm is developed and implemented. The details of the theoretical discussion as published by Kovesi is not repeated here. For that information, references [113, 114] is referred.

A.1 The Idea

The phase congruency algorithm is proposed by Kovesi in [113, 123]. The idea of phase congruency evolved from the top to the bottom in Figure A.1. The phase congruency is defined as the ratio of local energy $E(x)$ and the sum of the amplitude of Fourier components A_n[1]. The local energy $E(x)$ can be calculated as $\sum_n A_n \cos\big(\phi_n(x) - \bar{\phi}(x)\big)$, i.e. the energy is proportional to the cosine of the deviation of phase angle $\phi_n(x)$ from the overall mean phase angle $\bar{\phi}(x)$. Such relation can be easily derived from the illustration in

[1] Here n refers to the n^{th} Fourier component.

Figure A.2.

$$\begin{tikzcd}
PC(x)=\dfrac{E(x)}{\sum_n A_n} \arrow[r] \arrow[d, dashed] & \dfrac{\sum_n A_n \cos(\phi_n(x)-\bar{\phi}(x))}{\sum_n A_n}=\dfrac{\sum_n A_n \Delta\Phi(x)}{\sum_n A_n} \\
PC(x)=\dfrac{E(x)}{\sum_n A_n+\varepsilon} \arrow[r] \arrow[d, dashed] & \varepsilon \quad \text{a small positive constant} \\
PC(x)=\dfrac{\lfloor E(x)-T\rfloor}{\sum_n A_n+\varepsilon} \arrow[r] \arrow[d, dashed] & T \quad \text{estimated noise influence} \\
PC(x)=\dfrac{W(x)\lfloor E(x)-T\rfloor}{\sum_n A_n+\varepsilon} \arrow[r] \arrow[d, dashed] & W(x) \ \text{weighting function} \\
PC(x)=\dfrac{\sum_n W(x)\lfloor A_n(x)\Delta\Phi_n(x)-T\rfloor}{\sum_n A_n(x)+\varepsilon} \arrow[d, dashed] & \text{phase congruency across scale n} \\
PC(x)=\dfrac{\sum_o\sum_n W_o(x)\lfloor A_{no}(x)\Delta\Phi_{no}(x)-T_o\rfloor}{\sum_o\sum_n A_n(x)+\varepsilon} & \begin{array}{l}\text{phase congruency extended to}\\ \text{two dimensions}\end{array}
\end{tikzcd}$$

Figure A.1: The development of phase congruency algorithm.

A small positive value ε is added to the denominator in case all the Fourier amplitudes are very small. The radius of the noise circle is determined by the value T. A weighting function $W(x)$ is constructed to devalue phase congruency at locations where the spread of filter responses is low.

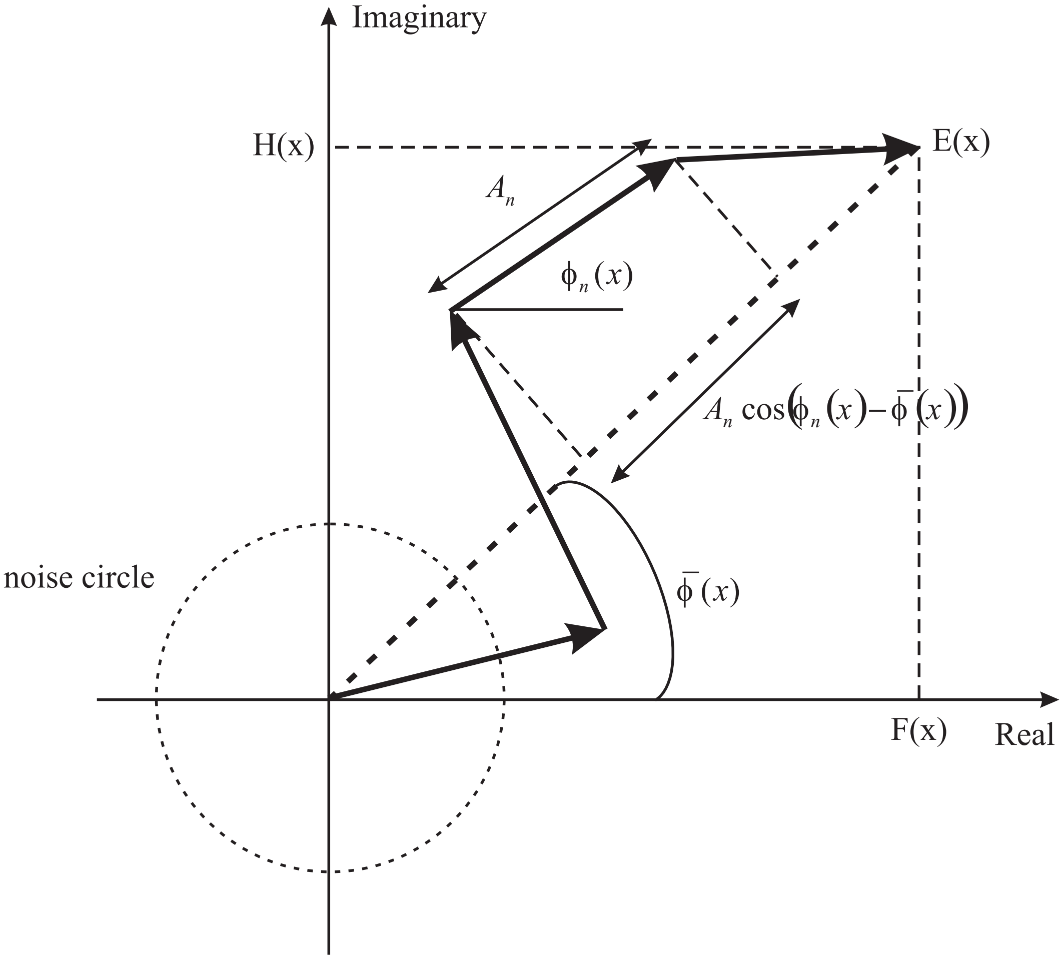


Figure A.2: Polar diagram showing the Fourier components at a location in the signal plotted head to tail (cf. Kovesi [3]).

By incorporating a more sensitive phase deviation measure $\Delta\Phi\,(x)$:

$$\Delta\Phi\,(x) = \cos\big(\phi_n(x) - \bar{\phi}(x)\big) - |\sin\big(\phi_n(x) - \bar{\phi}(x)\big)| \tag{A.1}$$

the phase congruency becomes:

$$PC(x) = \frac{\sum_n W(x) \lfloor A_n(x) \Delta\Phi_n(x) - T \rfloor}{\sum_n A_n(x) + \varepsilon} \tag{A.2}$$

Here, n indicates different scales. Two-dimensional phase congruency is calculated by combining data over several orientations. There is:

$$PC\left(x\right) = \frac{\sum_o \sum_n W_o\left(x\right) \left\lfloor A_n\left(x\right) \left(\cos\left(\phi_n\left(x\right) - \bar{\phi}\left(x\right)\right) - \left|\sin\left(\phi_n\left(x\right) - \bar{\phi}\left(x\right)\right)\right|\right) - T_o \right\rfloor}{\sum_o \sum_n A_n\left(x\right) + \varepsilon} \tag{A.3}$$

where o indicates the predefined orientations.

A.2 The Implementation

The detailed implementation is presented in Figure A.3 and A.4 below. The logarithmic Gabor function based wavelets are used to calculate phase congruency. Let M_n^e and M_n^o be the even-symmetric and odd-symmetric wavelets at a scale n. The 1D filers at four scales are shown in Figure A.5. The even- and odd-symmetric wavelets along six are given in Figure A.6 and A.7 respectively.

The convolution results of the input image $I\left(x\right)$ with quadrature pairs of filters at scale n are $e_n\left(x\right) = I\left(x\right) * M_n^e$ and $o_n\left(x\right) = I\left(x\right) * M_n^o$, which consist of the basic components to calculate $PC\left(x\right)$. There are:

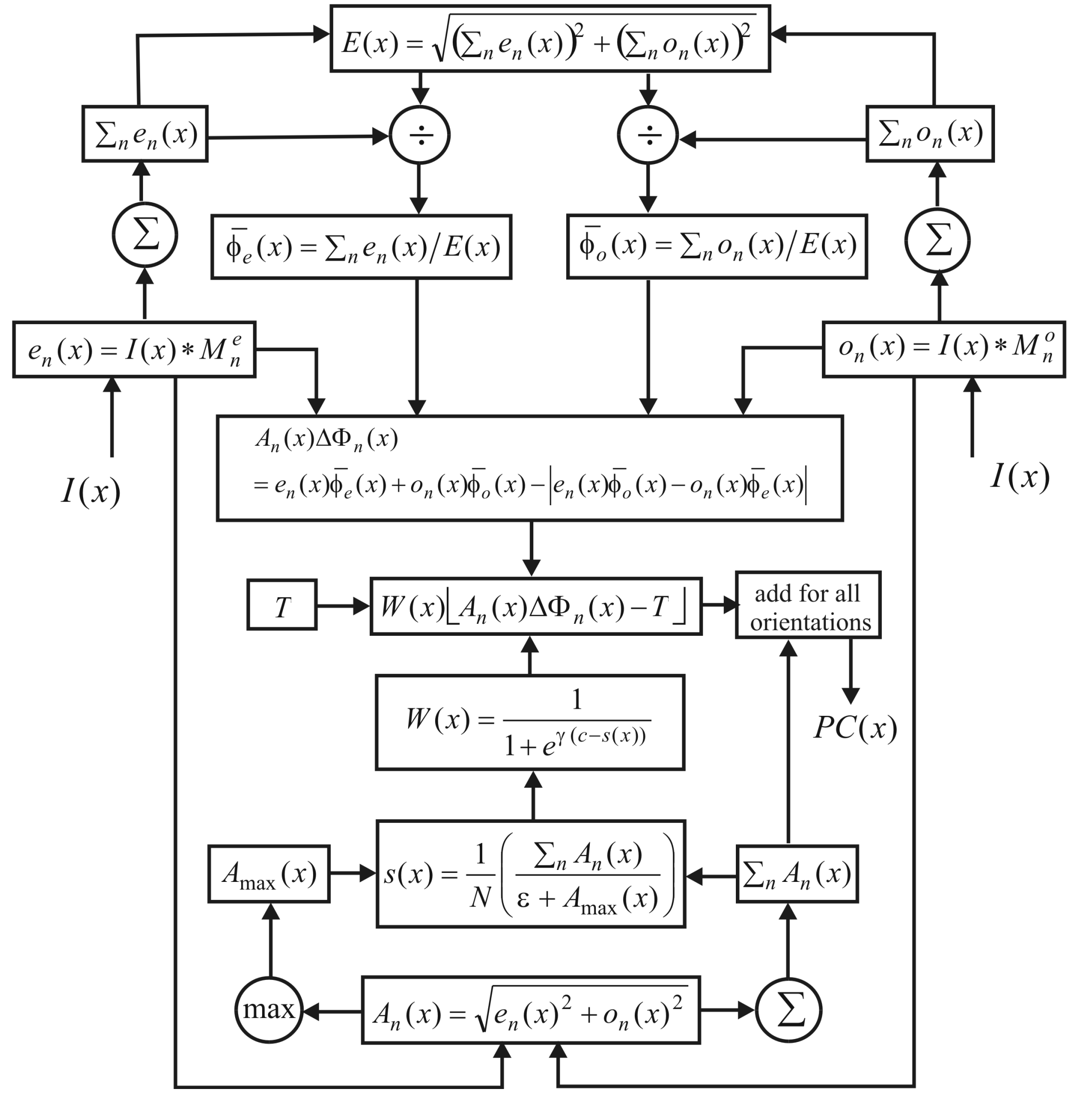

$E(x)=\sqrt{\left(\sum_n e_n(x)\right)^2+\left(\sum_n o_n(x)\right)^2}$
$\sum_n e_n(x)$
$\div$
$\div$
$\sum_n o_n(x)$
$\sum$
$\bar{\phi}_e(x)=\sum_n e_n(x)/E(x)$
$\bar{\phi}_o(x)=\sum_n o_n(x)/E(x)$
$\sum$
$e_n(x)=I(x)*M_n^e$
$o_n(x)=I(x)*M_n^o$
$A_n(x)\Delta\Phi_n(x)$
$=e_n(x)\bar{\phi}_e(x)+o_n(x)\bar{\phi}_o(x)-\left|e_n(x)\bar{\phi}_o(x)-o_n(x)\bar{\phi}_e(x)\right|$
$I(x)$
$I(x)$
T
$W(x)\lfloor A_n(x)\Delta\Phi_n(x)-T\rfloor$
add for all orientations
$PC(x)$
$W(x)=\frac{1}{1+e^{\gamma(c-s(x))}}$
$A_{\max}(x)$
$s(x)=\frac{1}{N}\left(\frac{\sum_n A_n(x)}{\varepsilon+A_{\max}(x)}\right)$
$\sum_n A_n(x)$
max
$A_n(x)=\sqrt{e_n(x)^2+o_n(x)^2}$
$\sum$

Figure A.3: The implementation of phase congruency algorithm. M_n^e and M_n^o denote the even-symmetric and odd-symmetric wavelet at this scale respectively.

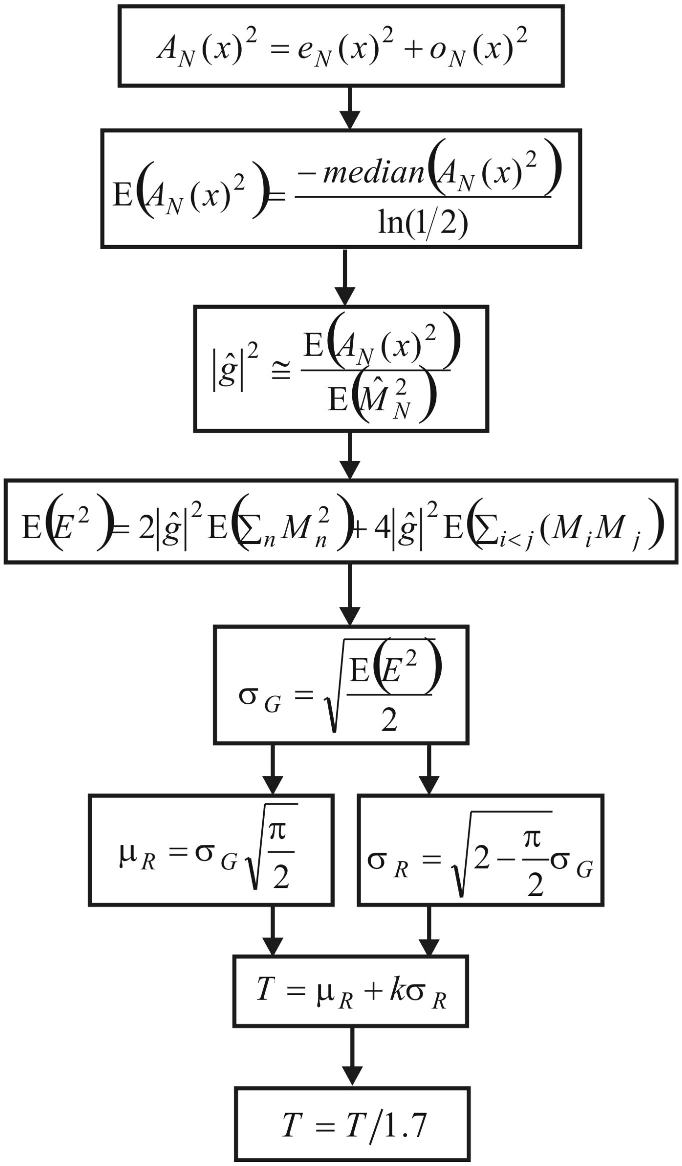

$A_N(x)^2 = e_N(x)^2 + o_N(x)^2$
$\mathrm{E}\left(A_N(x)^2\right) = \frac{-median\left(A_N(x)^2\right)}{\ln(1/2)}$
$|\hat{g}|^2 \cong \frac{\mathrm{E}\left(A_N(x)^2\right)}{\mathrm{E}\left(\hat{M}_N^2\right)}$
$\mathrm{E}\left(E^2\right) = 2|\hat{g}|^2 \mathrm{E}\left(\sum_n M_n^2\right) + 4|\hat{g}|^2 \mathrm{E}\left(\sum_{i<j}(M_i M_j\right)$
$\sigma_G = \sqrt{\frac{\mathrm{E}\left(E^2\right)}{2}}$
$\mu_R = \sigma_G \sqrt{\frac{\pi}{2}}$
$\sigma_R = \sqrt{2 - \frac{\pi}{2}\sigma_G}$
$T = \mu_R + k\sigma_R$
$T = T/1.7$

Figure A.4: The computation of noise compensation parameter T.

$$\begin{aligned}
\left[e_n(x), o_n(x)\right] &= \left[I(x) * M_n^e, I(x) * M_n^o\right] && \text{(A.4)}\\
\left[\bar{\phi}_e\left(x\right), \bar{\phi}_o\left(x\right)\right] &= \left[\cos\bar{\phi}(x), \sin\bar{\phi}(x)\right] && \text{(A.5)}\\
\cos\bar{\phi}(x) &= \frac{F(x)}{\sqrt{F(x)^2 + H(x)^2}} && \text{(A.6)}\\
\sin\bar{\phi}(x) &= \frac{H(x)}{\sqrt{F(x)^2 + H(x)^2}} && \text{(A.7)}
\end{aligned}$$

Referring to Figure A.2, the estimates of $F(x)$ and $H(x)$ can be obtained from:

$$\begin{aligned}
F\left(x\right) &= \sum\nolimits_n e_n\left(x\right) && \text{(A.8)}\\
H\left(x\right) &= \sum\nolimits_n o_n\left(x\right) && \text{(A.9)}\\
\sum\nolimits_n A_n(x) &= \sum\nolimits_n \sqrt{e_n(x)^2 + o_n(x)^2} && \text{(A.10)}
\end{aligned}$$

Using dot and cross products the following quantities can be formed:

$$\begin{aligned}
A_n(x)\cos\left(\phi_n(x) - \bar{\phi}(x)\right) &= e_n(x)\bar{\phi}_e(x) + o_n(x)\bar{\phi}_o(x) && \text{(A.11)}\\
A_n(x) \mid \sin\left(\phi_n(x) - \bar{\phi}(x)\right) \mid &= \mid e_n(x)\bar{\phi}_e(x) - o_n(x)\bar{\phi}_o(x) \mid && \text{(A.12)}
\end{aligned}$$

Therefore, there is:

$$\begin{aligned}
A_n(x)\Delta\Phi_n(x) &= A_n(x)\left(\cos\left(\phi_n(x) - \bar{\phi}(x)\right) - \mid \sin\left(\phi_n(x) - \bar{\phi}(x)\right) \mid\right) && \text{(A.13)}\\
&= e_n(x)\bar{\phi}_e(x) + o_n(x)\bar{\phi}_o(x) - \mid e_n(x)\bar{\phi}_e(x) - o_n(x)\bar{\phi}_o(x) \mid && \text{(A.14)}
\end{aligned}$$

The weighting function that devalues phase congruency is constructed from a measure of filter response spread $s(x)$ suggested by Kovesi [113]:

$$s\left(x\right) = \frac{1}{N}\left(\frac{\sum_{n} A_{n}\left(x\right)}{\varepsilon + A_{\max}\left(x\right)}\right) \tag{A.15}$$

$$W\left(x\right) = \frac{1}{1 + e^{\gamma\left(c - s\left(x\right)\right)}} \tag{A.16}$$

where $A_{max}(x)$ is the amplitude of the filter having the maximum response at x and ε is a small positive value to avoid division by zero. Value c and γ are the cut-off value of filter response spread and gain factor controlling the sharpness of the cut-off.

The compensation for the influence of noise T is estimated empirically for the mean noise response plus some multiple, k, of σ_R, i.e. $T = \mu_R + k\sigma_R$. However, such calculate has not been finally optimized yet. As shown in Figure A.3, the final phase congruency map can be obtained by combing the results from predefined orientations.

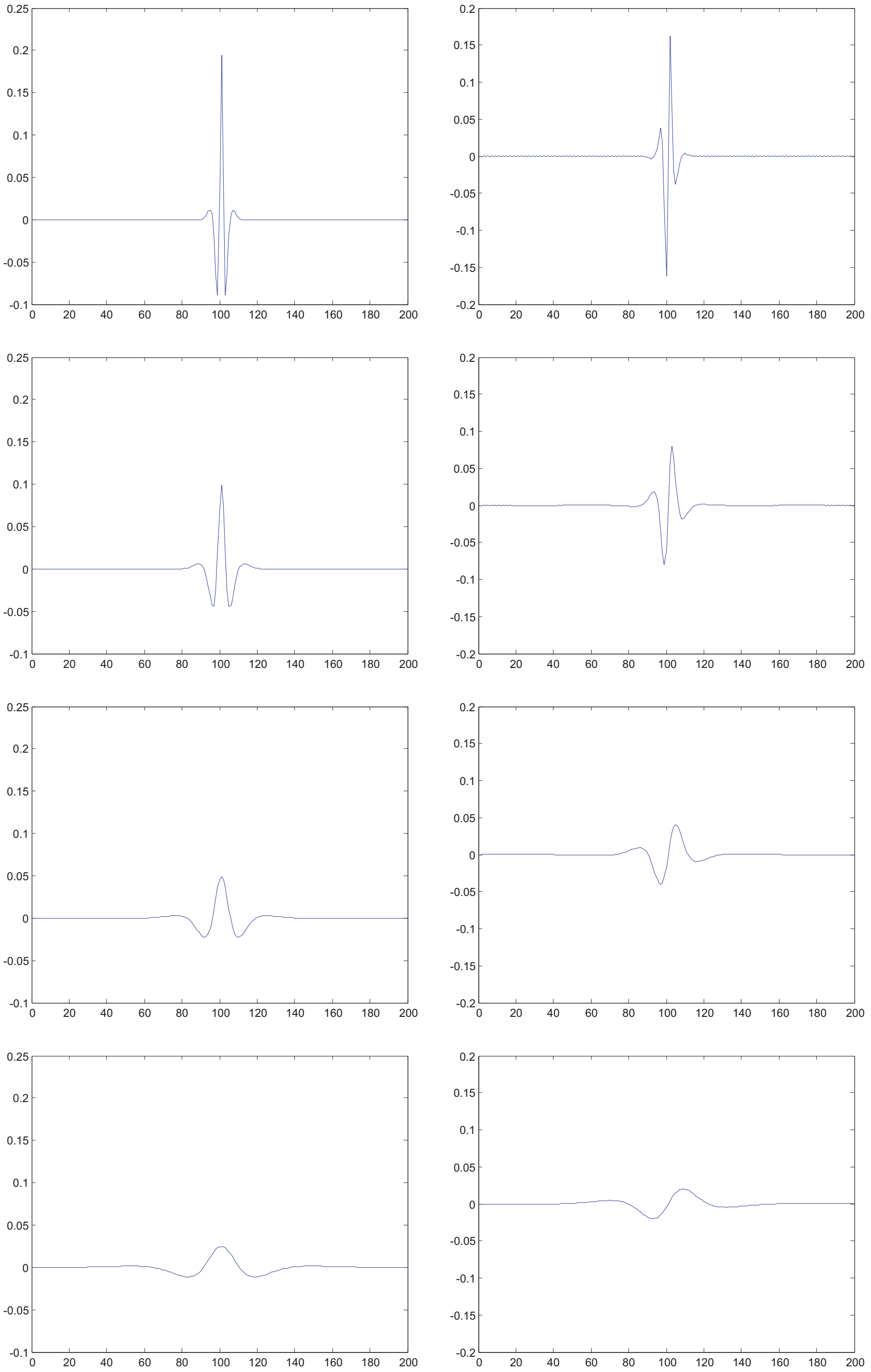


Figure A.5: The 1D log Gabor filters (left: even filter; right: odd filter; top to bottom: scale from 1 to 4).

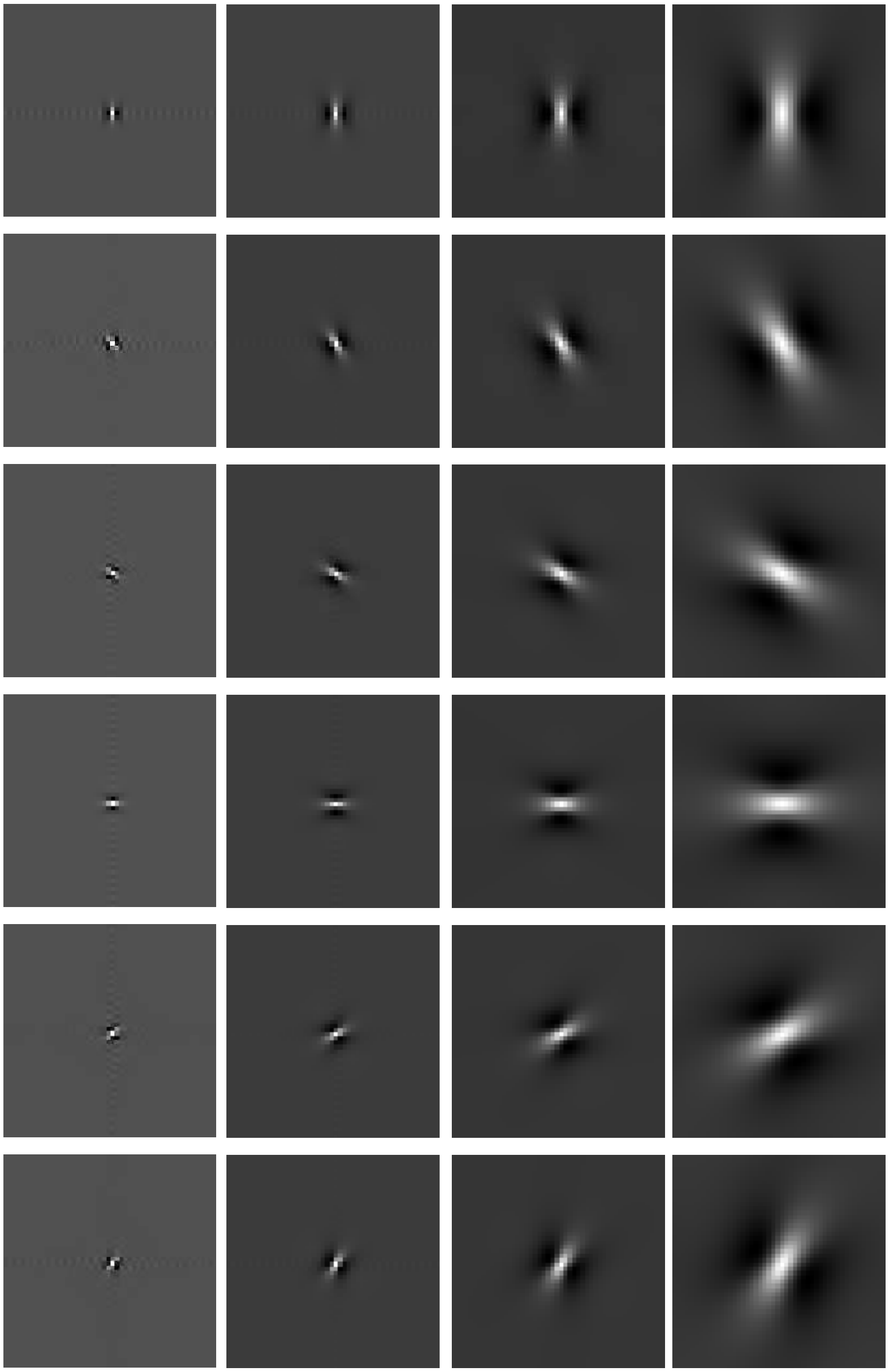


Figure A.6: The even filters (left to right: scale from 1 to 4; top to bottom: orientation from 1 to 6).

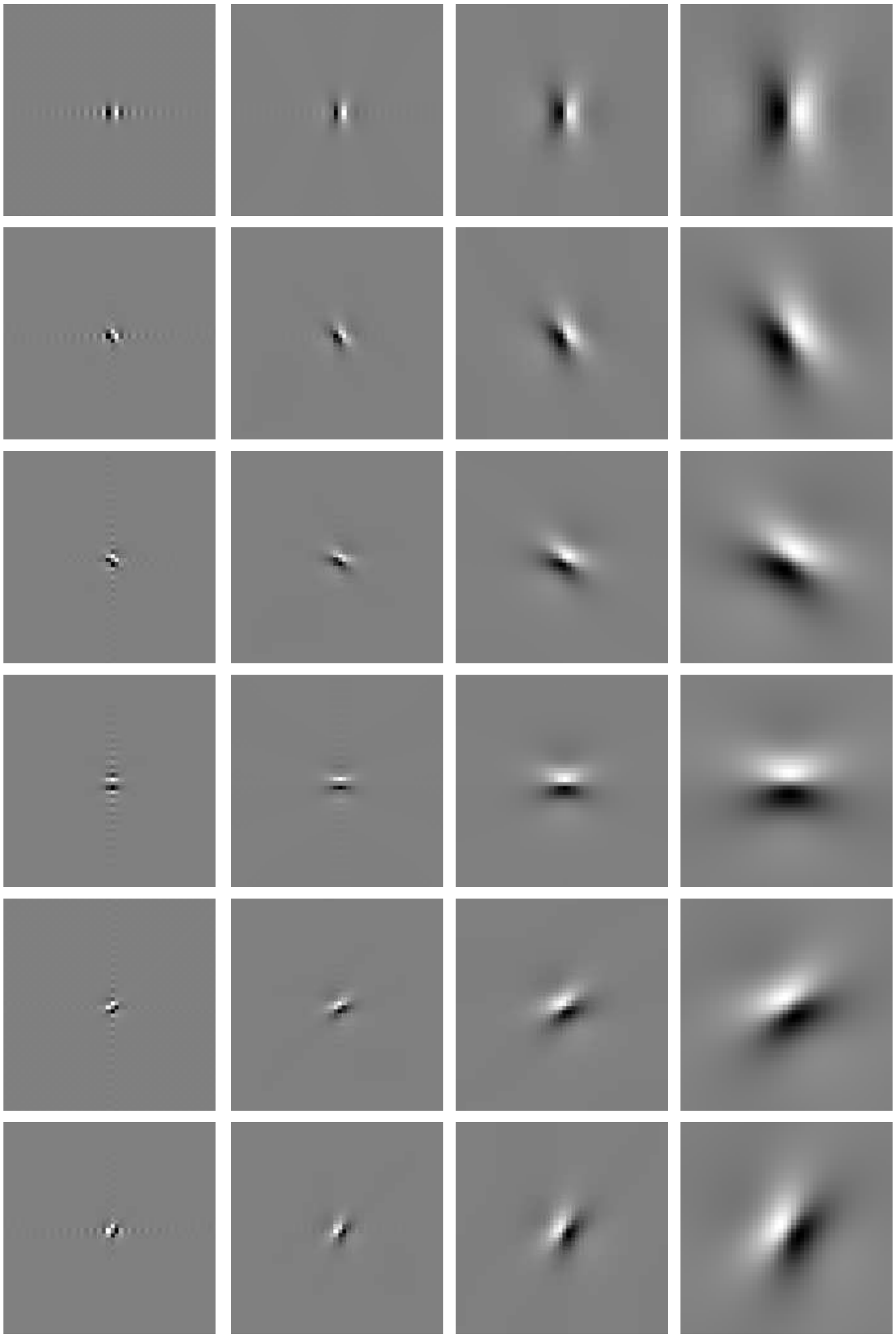


Figure A.7: The odd filters (left to right: scale from 1 to 4; top to bottom: orientation from 1 to 6).

Appendix B

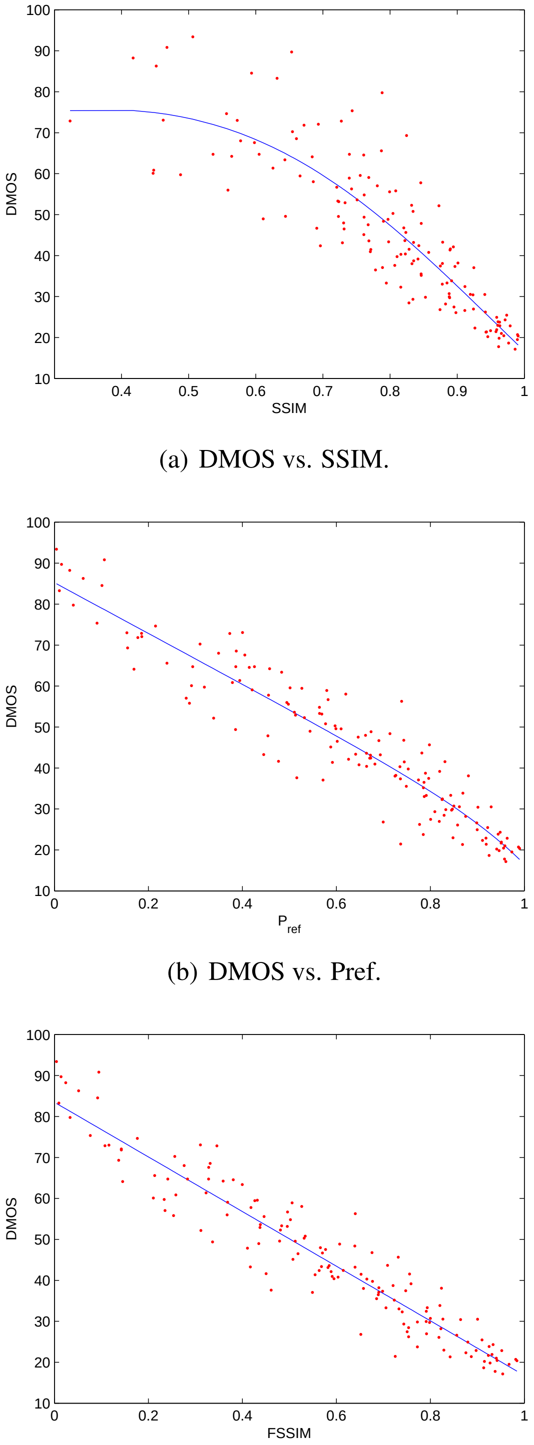
The Experiments with SSIM

The problem of the image structural similarity measurement (SSIM) is with the severely blurred images. Herein, the blurring operation is referred to the filtering with a circular-symmetric 2-D Gaussian kernel of certain standard deviation σ. We use the image database prepared by the Laboratory for Image and Video Engineering at the the University of Texas at Austin [124]. Twenty-nine images were used to create the database. The perceptual quality roughly covered the entire quality range. The raw scores obtained from human subject experiments were converted to the difference mean opinion score (DMOS) value or each distorted image [125].

The metrics proposed in Chapter 5, i.e. P_{ref} and FSSIM, are applied to the Gaussian-blurred images in the database and compared with the DMOS values. The relationship is modeled with a five-parameter logistic regression. The predicted values are compared with the experimental results and evaluation with root mean square error (RMSE), Spearman rank order correlation (SROC), outlier ratio, and cross-correlation. The results are listed in Table B.1. We can see both P_{ref} and FSSIM are better than SSIM in terms of the selected metrics.

Table B.1: The comparison of the predicted and experimental results.

	SSIM	P_{ref}	FSSIM
RMSE	8.9620	6.1374	5.6808
SROC	0.8945	0.9370	0.9508
Outlier Ratio	56.5517	44.8276	44.1379
CORR	0.8744	0.9432	0.9515



(a) DMOS vs. SSIM.

(b) DMOS vs. Pref.

(c) DMOS vs. FSSIM.

Figure B.1: The solid curves are obtained by five-parameter logistic regression.

Appendix C

Image Acknowledgements

The experiments described in Chapter 2 were validated with video sequences recorded for the MONNET project which was funded by Precarn Inc. and leaded by the Computer Vision and Systems Laboratory at Laval University.

The Signal Processing and Communication Laboratory at Lehigh University provides the multi-sensor images used for the study of fusion performance assessment and concealed weapon detection. Multi-sensor images used in Chapter 3 are obtained from the "image fusion website" at `http://www.imagefusion.org`.

The images "gold hill" and "Lena" used in Chapter 5 are provided by Dr. Z. Wang at the University of Texas (Austin).

The images used in the Appendix B are prepared by Dr. H. R. Sheikh at the University of Texas (Austin).

Bibliography

[1] "`http://www.siliconfareast.com/emspectrum.htm`," August 2007.

[2] "`http://www.physics.gatech.edu/academics/tutorial/phys2122/Chapter%2034/ir.htm`," August 2007.

[3] P. Kovesi, "Phase Congruency: A Low-level Image Invariant," *Psychological Research*, vol. 64, pp. 134–148, 2000.

[4] I. Moria and J. P. Heather, "Review of Image Fusion Technology in 2005," in *Proceedings of SPIE*, G. Raymond Peacock et al., Ed., Bellingham, WA, 2005, vol. 5782, pp. 29–45.

[5] P. J. Burt and R. J. Kolczynski, "Enhanced Image Capture through Fusion," in *Proceedings of International Conference on Image Processing*, 1993, pp. 248–251.

[6] M. A. Slamani, L. Ramac, M. Uner, P. Varshney, D. D. Weiner, M. Alford, D. Derris, and V. Vannicola, "Enhancement and Fusion of Data for Concealed Weapons Detection," in *Proceedings of SPIE*, 1997, vol. 3068, pp. 20–25.

[7] M. K. Uner, L. C. Ramac, P. K. Varshney, and M. Alford, "Concealed Weapon Detection: An Image Fusion Approach," in *Proceedings of SPIE*, 1996, vol. 2942, pp. 123–132.

[8] P. K. Varshney, L. Ramac, M. A. Slamani, M. G. Alford, and D. Ferris, “Fusion and Partitioning of Data for the Detection of Concealed Weapons,” in *Proceedings of the International Conference on Multisource-Multisensor Information Fusion*, 1998.

[9] P. K. Varshney, H. Chen, and M. Uner, “Registration and Fusion of Infrared and Millimetre Wave Images for Concealed Weapon Detection,” in *Proceedings of International Conference on Image Processing*, 1999, vol. 13, pp. 532–536.

[10] J. K. Aggarwal, *Multisensor Fusion for Computer Vision*, vol. 99 of *NATO ASI Series F: Computer and Systems Science*, 1993.

[11] Z. Xue, R.S. Blum, and Y. Li, “Fusion of Visual and IR Images for Concealed Weapon Detection,” in *Proceedings of ISIF 2002*, 2002, pp. 1198–1205.

[12] G. L. Foresti and L. Snidaro, “A distributed Sensor Network for Video Surveillance of Outdoors,” in *Multisensor Surveillance Systems*, G. L. Foresti, C. S. Regazzoni, and P. K. Varshney, Eds. Kluwer Academic Publishers, 2002.

[13] G. K. Matsopoulos, S. Marshall, and J. N. H. Brunt, “Multiresolution Morphological Fusion of MR and CT Images of the Human Brain,” *IEE Proceedings on Vision, Image and Signal Processing*, vol. 141, no. 3, pp. 137–142, 1994.

[14] I. Koren, A. Laine, and F. Taylor, “Enhancement via Fusion of Mammographic Features,” in *Proceedings of International Conference on Image Processing*, 1998, pp. 722–726.

[15] C. Pohl and J. L. Van Genderen, “Multi-sensor Image Fusion in Remote Sensing: Concepts, Methods and Applications,” *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823–854, 1998.

[16] X. E. Gros, Z. Liu, K. Tsukada, and K. Hanasaki, “Experimenting with Pixel-Level NDT Data Fusion Techniques,” *IEEE Transaction on Instrumentation and Measurement*, vol. 49, no. 5, pp. 1083–1090, October 2000.

[17] X. P. V. Maldague, *Theory and Practice of Infrared Technology for Nondestructive Testing*, Wiley Series in Microwave and Optical Engineering. John Wiley and Sons, Inc., 2001.

[18] G. B. Rybicki and A. P. Lightman, *Radiative Processes in Astrophysics*, John Wiley and Sons, New York, USA, 1979.

[19] Rick S. Blum and Zheng Liu, Eds., *Multi-sensor Image Fusion and Its Applications*, Signal Processing and Communications. Taylor and Francis, 2005.

[20] Y. Zheng, E. A. Essock, and B. C. Hansen, “Advanced Discrete Wavelet Transform Fusion Algorithm and Its Optimization by Using the Metric of Image Quality Index,” *Optical Engineering*, vol. 44, no. 3, pp. 037003 (12 Pages), March 2005.

[21] L. Wald, “Some Terms of Reference in Data Fusion,” *IEEE Transactions on Geosciences and Remote Sensing*, vol. 37, no. 3, pp. 1190–1193, 1999.

[22] Z. Liu, D. S. Forsyth, J. P. Komorowski, and K. Hanasaki an T. Kirubarajan, “Survey: State of the Art in NDE Data Fusion,” *IEEE Transaction on Instrumentation and Measurement*, vol. 56, no. 5, October 2007.

[23] B. Zitova and J Flusser, “Image Registration Methods: A Survey,” *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[24] H. Li and Y.-T. Zhou, “Automatic Visual/IR Image Registration,” *Optical Engineering*, vol. 35, no. 2, pp. 391–400, February 1996.

[25] H. Li, B. S. Manjunath, and S. K. Mitra, “A Contour-Based Approach to Multisensor Image Registration,” *IEEE Transactions on Image Processing*, vol. 4, no. 3, pp. 320–334, March 1995.

[26] E. Coiras, J. Santamaria, and C. Miravet, “Segment-based Registration Technique for Visual-Infrared Images,” *Optical Engineering*, vol. 39, no. 1, pp. 282–289, January 2000.

[27] J. Han and B. Bhanu, “Detecting Moving Humans Using Color and Infrared Video,” in *Proceedings of IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2003, pp. 228–233.

[28] G. Ye, J. Wei, M. R. Rickering, M. R. Frater, and J. F. Arnold, “Simultaneous Tracking and Registration in A Multisensor Surveillance System,” in *Proceedings of International Conference on Image Processing*, Sept 14-17 2003, vol. 1, pp. 933–936.

[29] Y. Keller and A. Averbuch, “Multisensor Image Registration via Implicite Similarity,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 794–801, May 2006.

[30] W. D. Jones, “Safer Driving in the Dead Night,” *IEEE Spectrum*, pp. 20–21, March 2006.

[31] J. A. Stark, “Adaptive Image Contrast Enhancement Using Generalizations of Histogram Equalization,” *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 889–896, May 2000.

[32] Z. Yu and C. Bajaj, "A Fast and Adaptive Method fo Image Contrast Enhancement," in *Proceedings of International Conference on Image Processing*, Oct 24-28 2004, pp. 1001–1004.

[33] R. S. Blum, Zhiyun Xue, and Zhong Zhang, "An Overview of Image Fusion," in *Multi-Sensor Image Fusion and Its Applications*, Rick S. Blum and Zheng liu, Eds., Chapter 1, pp. 1–35. Taylor and Francis, 2005.

[34] L. Tao, H. Ngo, M. Zhang, A. Livingston, and V. Asari, "A Multi-Sensor Image Fusion and Enhancement System for Assisting Drivers in Poor Lighting Conditions," in *Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop*, 2005.

[35] B. A. Klock, "Interface and Usability Assessment of Imaging Systems," *IEEE AESS Systems Magazine*, pp. 11–12, March 2003.

[36] R. W. McMillan, Jr. O. Milton, M. C. Hetzler, R. S. Hyde, and W. R. Owens, "Detection of Concealed Weapons Using Far-Infrared Bolometer Arrays," in *Conference Digest on 25th Infrared and Millimeter Waves*, Sept 12-15 2000, pp. 259–260.

[37] N. G. Paulter, *Guide to the Technologies of Concealed Weapon and Contraband Imaging and Detectionm (NIJ Guide 602-00)*, *U.S.* Department of Justice, Office of Justice Program, National Insitute of Justice, February 2001.

[38] H. M. Chen, S. Lee, R. M. Rao, M. A. Slamani, and P. K. Varshney, "Imaging for Concealed Weapon Detection," *IEEE Singal Processing Magazine*, vol. 22, no. 2, pp. 52–61, March 2005.

[39] Y. Wang and B. Lohmann, "Multisensor Image Fusion: Concept, Method and Applications," Tech. Rep., Institut für Automatisierungstechnik, Universität Bremen, German, December 2000.

[40] Z. Wang and A. C. Bovik, “A Universal Image Quality Index,” *IEEE Signal Processing Letters*, vol. 9, no. 8, pp. 81–84, 2002.

[41] G. Qu, D. Zhang, and P. Yan, “Information Measure for Performance of Image Fusion,” *Electronics Letters*, vol. 38, no. 7, pp. 313–315, 2002.

[42] C. S. Xydeas and V. Petrovic, “Objective Image Fusion Performance Measure,” *Electronics Letters*, vol. 36, no. 4, pp. 308–309, 2000.

[43] C. S. Xydeas and V. Petrovic, “Objective Pixel-Level Image Fusion Performance Measure,” in *Proceedings of SPIE*, 2000, vol. 4051, pp. 89–98.

[44] G. Piella and H. Heijmans, “A New Quality Metric for Image Fusion,” in *Proceedings of International Conference on Image Processing*, Bacelona, 2003.

[45] G. Piella, “New Quality Measures for Image Fusion,” in *Proceedings of International Conference on Information Fusion*, Stockholm, Sweden, 2004.

[46] S. Krotosky and M. Trivedi, “Multimodal Stereo Image Registration for Pedestrian Detection,” in *Proceedings of Intelligent Transportation Systems*, Toronto, Canada, September 2006, pp. 109–114.

[47] A. El-Maadi, V. Gregorie, L. St-Laurent, H. Torresan, B. Turgeon, D. Prevost, P. Hebert, D. Laurendeau, B. Ricard, and X. Maldague, “Visible and infrared Imagery for Surveillance Applications: Software and Hardware Considerations,” *International Journal on Quantitative Infrared Thermography*, vol. 4, no. 1, pp. 25–40, 2007.

[48] H. Torresan, B. Turgeon, C. Ibarra-Castanedo, P. Hébert, and X. Maldague, “Advanced Surveillance System: Combining Video and Thermal Imagery for Pedestrian

Detection," in *Proc. of SPIE, Thermosense XXVI*, G. Raymond Peacock Douglas D. Burleigh, K. Elliott Cramer, Ed., April 2004, vol. 5405 of *SPIE*, pp. 506–515.

[49] K. Yasuda, T. Naemura, and H. Harashima, "Thermo-Key Human Region Segmentation from Video," *Computer Graphics and Applications*, vol. 24, no. 1, pp. 26–30, 2004.

[50] F. Xu, X. Liu, and K. Fujimura, "Pedestrian Detection and Tracking with Night Vision," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 63–71, March 2005.

[51] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," *IEEE Transaction on Medical Imaging*, vol. 16, no. 2, pp. 187–198, April 1997.

[52] H. M. Chen, P. K. Varshney, and M. A. Slamani, "On Registration of Regions of Interest (ROI) in Video Sequences," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, Los Alamitos, CA, USA, 2003, pp. 313–318.

[53] 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*, vol. 13, no. 4, pp. 600–612, April 2004.

[54] "`http://en.wikipedia.org/wiki/Infrared_camera`," August 2007.

[55] A. Toet, *Fusion of Image from Different Electro-Optical Sensing Modalities for Surveillance and Navigation Tasks*, Chapter 7, pp. 237–264, Taylor and Francis, 2005.

[56] E. J. Bender, C. E. Reese, and G. S. van der Wal, "Comparison of Additive Image Fusion vs. Feature-level Image Fusion Techniuqes for Enhanced Night Driving," in *Proceedings of SPIE*, C. Bruce Johnson, Divyendu Sinha, and Philip A. Lapante, Eds., 2003, vol. 4796, pp. 140–151.

[57] J. Yang and R. S. Blum, *A Statistical Signal Processing Approachto Image Fusion Using Hidden Markov Models*, Chapter 8, pp. 265–287, Signal Processing and Communications. Taylor and Francis, 2005.

[58] J. Yang and R. S. Blum, "A Statistical Signal Processing Approach to Image Fusion for Concealed Weapon Detection," in *Proceedings of International Conference on Image Processing*, 2002, vol. 1, pp. 513–516.

[59] D. A. Fay, A. M. Waxman, M. Aguilar, D. B. Ireland, J. P. Racamato, W. D. Ross, W. W. Streilein, and M. I. Braun, "Fusion of Multi-Sensor Imagery for Night Vision: Color Visualization, Target Learning and Search," in *Proceedings of ISIF 2000*, 2000, pp. TuD3: 3–10.

[60] Z. Xue and R. S. Blum, "Concealed Weapon Detection Using Color Image Fusion," in *Proceedings of 6th International Conference of Information Fusion*, 2003, vol. 1, pp. 622–627.

[61] A. Toet, J. K. IJspeert, A. M. Waxman, and M. Aguilar, "Fusion of Visual and Thermal Imagery Improves Situational Awareness," *Displays*, vol. 18, pp. 85–95, 1997.

[62] G. Piella, "A General Framework for Multiresolution Image Fusion: from Pixels to Regions," *Information Fusion*, vol. 4, no. 4, pp. 259–280, December 2003.

[63] Z. Zhang, *Investigations of Image Fusion*, Ph.D. thesis, Lehigh University, 1999.

[64] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid Methods in Image Processing," *RCA Engineer*, vol. 29, no. 6, pp. 33–41, 1984.

[65] A. Teot, "Image Fusion by a Ratio of Low-pass Pyramid," *Pattern Recognition Letters*, vol. 9, pp. 245–253, 1989.

[66] A. Toet, "Multiscale Contrast Enhancement with Application to Image Fusion," *Optical Engineering*, vol. 31, no. 5, pp. 1026–1031, May 1992.

[67] T. A. Wilson, S. K. Rogers, and L. R. Myers, "Perceptual-based Hyperspectral Image fusion Using Multiresolution Analysis," *Optical Engineering*, vol. 34, no. 11, pp. 3154–3164, November 1995.

[68] T. A. Wilson, S. K. Rogers, and M. Kabrisky, "Perceptual-based Image Fusion for Hyperspectral Data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1007–1017, July 1997.

[69] Z. Liu, K. Tsukada, K. Hanasaki, Y. K. Ho, and Y. P. Dai, "Image Fusion by Using Steerable Pyramid," *Pattern Recognition Letters*, vol. 22, pp. 929–939, 2001.

[70] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor Image Fusion Using the Wavelet Transform," *Graphical Models and Image Processing*, vol. 57, no. 3, pp. 235–245, 1995.

[71] I. Koren, A. Laine, and F. Taylor, "Image Fusion Using Steerable Dyadic Wavelet Transform," in *Proceedings of International Conference on Image Processing*, 1995, pp. 232–235.

[72] O. Rockinger, "Pixel level fusion of image sequences using wavelet frames," in *Proc. 16th Leeds Annual Statistical Research Workshop*. 1996, pp. 149–154, Leeds University Press.

[73] O. Rockinger, “Image Sequence Fusion Using a Shift-invariant Wavelet Transform,” in *Proceedings of International Conference on Image Processing*, 1997, vol. 3, pp. 288–301.

[74] O. Rockinger and T. Fechner, “Pixel-level Image Fusion: The Case of Image Sequences,” in *SPIE*, 1998, vol. 3374, pp. 378–388.

[75] Z. Zhang and R. S. Blum, “Image Fusion for a Digital Camera Application,” in *Proceedings of 32nd Asilomar Conference on Signals Systems, and Computers*, Monterey, CA, 1998, pp. 603–607.

[76] T. Pu and G. Q. Ni, “Contrast-based Image Fusion Using Discrete Wavelet Transform,” *Optical Engineering*, vol. 39, no. 8, pp. 2075–2082, August 2000.

[77] S. Nikolov, P. Hill, D. Bull, and N. Canagarajah, “Wavelets for Image Fusion,” in *Wavelets in signal and image analysis, Computational Imaging and Vision Series*, A. Petrosian and F. Meyer, Eds., pp. 213–244. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.

[78] H. Wang, J. Peng, and W. Wu, “Fusion Algorithm for Multisensor Images based on Discrete Multiwavelet Transform,” *IEE Proceedings on Vision, Image Signal Process*, vol. 149, no. 5, pp. 283–289, October 2002.

[79] “`http://www.imagefusion.org`,” August 2007.

[80] V. Petrovic and C. Xydeas, “Multiresolution image fusion using cross band feature selection,” in *Proceedings of SPIE*, 1999, vol. 3719, pp. 319–326.

[81] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2nd Edition, January 2002.

[82] L. Tao and V. K. Asari, “Adaptive and Integrated Neighborhood-dependent Approach for Nonlinear Enhancement of Color Images,” *Journal of Electronic Imaging*, vol. 14, no. 4, pp. 1–14, Oct-Dec 2005.

[83] L. Tao and V. K. Asari, “An Efficient Illuminance-Reflectance Nonlinear Video Stream Enhancement Model,” in *Proceedings of SPIE*, N. Kehtarnavaz, Ed., 2006, vol. 6063 of *Real-Time Image Processing*.

[84] L. Tao, R. Tompkins, and V. K. Asari, “An Illuminance-Reflectance Nonlinear Video Enhancement Model for Homeland Security Applications,” in *Proceedings of 34th Applied Imagery and Pattern Recognition Workshop*, 2005.

[85] E. P. Siomoncelli, W. T. Freeman, E. H. Adelson, and D.J. Heege, “Shiftable Multiscale Transform,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.

[86] E.P. Siomoncelli and W.T. Freeman, “The Steerable Pyramid: A flexible Architecture for Multi-scale Derivative Computation,” in *Proceedings of International Conference on Image Processing*, Washington DC, 1995, pp. 444–447.

[87] “`http://www.cns.nyu.edu/~eero/STEERPYR`,” August 2007.

[88] A Karasaridis and E Simoncelli, “A Filter Design Technique for Steerable Pyramid Image Transforms,” in *Proceedings of International Conference on Acoustics Speech and Signal Processing*, Atlanta GA, May 1996, vol. 4, pp. 2387–2390.

[89] C. S. Bouganis, P.Y.K. Cheung, J. Ng, and A. A. Bharath, *A Steerable Complex Wavelet Construction and Its Implementation on FPGA*, vol. 3203 of *Lecture Notes in Computer Science*, pp. 394–403, 2004.

[90] Z. Liu, Z. Xue, R. S. Blum, and R. Laganiere, “Concealed Weapon Detection and Visualization in a Synthesized Image,” *Pattern Analysis and Applications*, vol. 8, no. 4, pp. 375–389, February 2006.

[91] P. Loftus, “Camera Detects Concealed Weapons,” *The Wall Street Journal (online)*, April 13 2005.

[92] M. A. Slamani, P. K.. Varshney, R. M. Rao, M. G. Alford, and D. Ferris, “Image Processing Tools for the Enhancement of Concealed Weapon Detection,” in *Proceedings of ICIP*, Kobe, Japan, Oct 24-28 1999, vol. 3, pp. 518–522.

[93] P. K. Varshney, H. Chen, and R. M. Rao, “On Siganl/Image Processing for Concealed Weapon Detection from Stand-off Range,” in *Proceedings of SPIE*, Theodore T. Saito, Ed., 2005, vol. 5781, pp. 93–97.

[94] M. A. Slamani, M. Alford, and D. Ferris, “Setting Thresholds in Infrared Images for the Detection of Concealed Weapons,” in *Proceedings of SPIE*, San Diego, California, July 1998, vol. 3460, pp. 630–639, SPIE.

[95] N. Otsu, “A Threshold Selection Method from Gray Level,” *IEEE Transaction on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, 1979.

[96] J. Yang and R. S. Blum, “Image Fusion Using the Expectation-Maximization Algorithm and a Hidden Markov Model,” in *Proceedings of IEEE Vehicular Technology Conference*, Los Angeles, September 2004, vol. 6, pp. 4563 – 4567.

[97] J. Yang and R. S. Blum, “A Region-based Image Fusion Method Using the Expectation-Maximization Algorithm,” in *Proceedings of Conference on Information Science and Systems*, 2006.

[98] R. S. Blum, Z. Xue, Z. Liu, and D. S. Forsyth, “Multisensor Concealed Weapon Detection by Using a Multiresolution Mosaic Approach,” in *Proceedings of IEEE Vehicular Technology Conference*, September 2004, vol. 7, pp. 4597–4601.

[99] R. O. Duda, P. E. Hart, and D.G. Strok, *Patten Classification*, Wiley Interscience, 2nd Edition, 2000.

[100] B. Balasko, J. Abonyi, and B. Feil, *Fuzzy Clustering and Data Analysis Toolbox*, Department of Process Engineering, University of Veszprem, Veszprem, Hungary.

[101] A. M. Bensaid, L. O. Hall, J. C. Bezdek, L. P. Clarke, M. L. Silbiger, J. A. Arrington, and R. F. Murtagh, “Validity-guided (Re)Clusting with Applications to Image Segmentation,” *IEEE Transactions on Fuzzy Systems*, vol. 4, pp. 112–123, May 1996.

[102] X. L. Xie and G. A. Beni, “Validity Measure for Fuzzy Clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841 – 847, 1991.

[103] U. Maulik and S. Bandyopadhyay, “Performance Evaluation of Some Clustering Algorithms and Validity Indices,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1650– 1654, 2002.

[104] C. T. Hsu and J. L. Wu, “Multiresolution Mosaic,” *IEEE Transactions on Consumer Electronics*, vol. 42, no. 4, pp. 981–990, November 1996.

[105] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, “Wallflower: Principles and practice of background maintenance,” in *Proceedings of International Conference on Computer Vision*, 1999, pp. 255–261.

[106] N. C. Currie, F. J. Demma, D. D. Ferris Jr., B. R. Kwasowsky, R. W. McMillan, and M. C. Wicks, "Infrared and Millimeter-Wave Sensors for Military Special Operations and Law Enforcement Applications," *International Journal of Infrared and Millimeter Waves*, vol. 17, no. 7, pp. 1117–1138, July 1996.

[107] G. Paolo, "Image Comparison Metrics : A Review," `citeseer.ist.psu.edu/lorenzetto98image.html`, May 1998.

[108] D. L. Wilson, A. J. Baddeley, and R. A. Owens, "A New Metric For Grey-Scale Image Comparison," *International Journal of Computer Vision*, vol. 24, no. 1, pp. 5–17, 1997.

[109] D. E. B. Lees and P. D. Henshaw, "Printed Circuit Board Inspection - A Novel Approach," in *Proceedings of SPIE (Automated Inspection and Measurement)*, 1986, pp. 164–173.

[110] G. P. Leronzetto and P. Kovesi, "A Phase Based Image Comparison Technique," in *Proceeding of DICTA*, Western Australia, 1999.

[111] P. Kovesi, *Invariant Measures of Image Features from Phase Information*, Ph.D. thesis, University of Western Australia, 1996.

[112] M. C. Morrone and R. A. Owens, "Feature Detection from Local Energy," *Pattern Recognition Letters*, vol. 6, pp. 303–313, 1987.

[113] P. Kovesi, "Image Features From Phase Congruency," *Videre: A Journal of Computer Vision Research*, vol. 1, no. 3, 1999.

[114] P. Kovesi, *Invariant Measures of Image Features from Phase Information*, Ph.D. thesis, University of Western Australia, May 1996.

[115] P. Kovesi, “Image Features from Phase Congruency,” Tech. Rep., University of Western Australia, 1995.

[116] P. Kovesi, “Invariant Measures of Image Feature from Phase Information, http://www.csse.uwa.edu.au/ pk/Research/research.html,” August 2007.

[117] J. Martin and J. L. Crowley, “Experimental Comparison of Correlation Techniques,” in *Proceedings of International Conference on Intelligent Autonomous Systems*, 1995.

[118] G.-H. Chen, C.-L. Yang, L.-M. Po, and S.-L. Xie, “Edge-based Structural Similarity for Image Quality Assessment,” in *Proceedings of ICASSP*, Toulouse, France, May 15-19 2006, vol. 2, pp. 933–936.

[119] T. Kubota, *Orientational Filters for Real-Time Computer Vision Problems*, Ph.D. thesis, Georgia Institute of Technology, 1995.

[120] Y. H. Cheung, H. W. Leong, K. C. Tsang, and E. Shi, “Implementation of Gabor-type Filters on Field Programmabel Gate Arrays,” in *Proceedings of Field-Programmable Technology*, December 2005, pp. 327–328.

[121] P. Kovei, “Phase Congruency Detects Corners and Edges,” in *Proceedings of the Australian Pattern Recognition Society Conference: DICTA 2003*, December 2003, pp. 309–318.

[122] D. V. Wekenm, M. Nachtegael, and E. E. Kerre, “Using Similarity Measures and Homogeneity for the Comparison of Images,” *Image and Visual Computing*, vol. 22, pp. 695–702, 2004.

[123] P. D. Kovesi, “MATLAB and Octave Functions for Computer Vision and Image Processing,” School of Computer Science & Software

Engineering, The University of Western Australia, Available from: <http://www.csse.uwa.edu.au/∼pk/research/matlabfns/>.

[124] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2," `http://live.ece.utexas.edu/research/quality`, August 2007.

[125] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.