

Investigations on Multi-Sensor Image System and Its Surveillance Applications

Zheng Liu

DISSERTATION.COM



Boca Raton

Investigations on Multi-Sensor Image System and Its Surveillance Applications

Copyright © 2007 Zheng Liu
All rights reserved.

Dissertation.com
Boca Raton, Florida
USA • 2008

ISBN-10: 1-59942-651-X
ISBN-13: 978-1-59942-651-8

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

© Copyright by Zheng Liu 2007

All Rights Reserved

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

Abstract	iii
Acknowledgments	vi
List of Tables	xi
List of Figures	xiii
1 Introduction	1
1.1 Motivation and Objective	1
1.1.1 The Statement of Problems	4
1.1.2 Objectives of the Research	6
1.2 Background and Significance	7
1.2.1 Multi-Sensor Image System	7
1.2.2 Implementation of Image Fusion for Surveillance Applications	11
1.2.3 Objective Assessment of the Fusion Performance	14
1.3 Organization of the Thesis	15
1.4 Contributions of the Thesis	17
2 Registration of Visual and Infrared Video Sequences	22
2.1 Introduction	22
2.2 Registration based on Frame Difference	25

2.2.1	Image Similarity Measurement	25
2.2.2	Silhouette Extraction	29
2.2.3	Parameter Estimation and Refinement	29
2.3	Experimental Results	34
2.4	Discussion	37
2.5	Conclusion	39
3	Context Enhancement through Infrared Vision	42
3.1	Introduction	42
3.2	Multiresolution Analysis (MRA) based Image Fusion: A Brief Review . . .	45
3.3	Enhancement and Fusion	52
3.3.1	Histogram-based Operations	52
3.3.2	Adaptive Enhancement	54
3.3.3	Pixel-level Image Fusion	58
3.4	A Modified Scheme	61
3.5	More Results	64
3.6	Discussion	66
3.7	Conclusion	76
4	Concealed Weapon Detection and Visualization in a Synthesized Image	78
4.1	Introduction	78
4.2	Problem Review	80
4.3	A Two-Step Scheme for Synthesizing a Composite Image	85
4.3.1	Concealed Weapon Detection	86
4.3.2	Embedding in a Visual Image	89
4.3.3	Result Assessment	92
4.4	Experimental Results	93

4.5	Discussion	100
4.6	Conclusion	107
5	The Use of Phase Congruency for Reference-based Assessment	108
5.1	Introduction	108
5.2	Typical Solutions	110
5.3	Image Feature from Phase Congruency	112
5.3.1	The Concept of Phase Congruency	112
5.3.2	Implementation of Phase Congruency Algorithm with the Logarithmic Gabor Filter	113
5.4	Reference-based Assessment for Image Fusion	114
5.4.1	Image Similarity Measurement	114
5.4.2	A Modified SSIM Scheme	117
5.5	Experimental Results	118
5.5.1	Experiments for Image Comparison	118
5.5.2	Experiments for Fusion Assessment	130
5.6	Discussion	130
5.7	Conclusion	136
6	Feature-based Metrics for Blind Assessment	137
6.1	Introduction	137
6.2	Blind Evaluation of Image Fusion	139
6.3	A Strategy for the Feature-based Evaluation	142
6.3.1	Principal Moments of Phase Congruency	142
6.3.2	Quality Metrics for Evaluating Image Fusion	144
6.4	Experimental Results	148
6.5	Discussion	151

6.6 Conclusion	165
7 Conclusions	167
Appendix A	
The Implementation of Phase Congruency Algorithm	171
A.1 The Idea	171
A.2 The Implementation	174
Appendix B	
The Experiments with SSIM	182
Appendix C	
Image Acknowledgements	185
Bibliography	186

List of Tables

1.1	The electromagnetic wavelength table [1]	2
2.1	The configuration parameters for Radiance PM IR camera.	34
2.2	The specifications for Pulnix TMC6700CL camera.	35
2.3	The registration parameters obtained by maximum MI.	35
3.1	Comparison of multiresolution image fusion schemes: image pyramid.	47
3.2	Comparison of multiresolution image fusion schemes: discrete wavelet.	48
4.1	The summary of the image fusion techniques for CWD.	83
4.2	Comparison of the Fuzzy k-means clustering results with different initial cluster number.	95
4.3	Comparison of clustering schemes.	100
5.1	The notation for equation (5.1)-(5.5).	111
5.2	Experimental results on image comparison (Gold Hill).	127
5.3	Experimental results on image comparison (Lena).	127
5.4	The standard deviation of the assessment results for image “gold hill” and “Lena”.	128
5.5	Evaluation of the fusion result of multi-focus image “laboratory”.	132
5.6	Evaluation of the fusion result of multi-focus image “books”.	132
5.7	Evaluation of the fusion result of multi-focus image “Japanese food”.	133
5.8	Evaluation of the fusion result of multi-focus image “Pepsi”.	133

5.9	Evaluation of the fusion result of multi-focus image “objects”.	134
5.10	The standard deviation of the assessment results for the fusion of multi-focus images.	134
6.1	Evaluation of the fusion results of multi-focus image “laboratory”.	148
6.2	Evaluation of the fusion results of multi-focus image “books”.	149
6.3	Evaluation of the fusion results of multi-focus image “Japanese food”.	149
6.4	Evaluation of the fusion results of multi-focus image “Pepsi”.	150
6.5	Evaluation of the fusion results of multi-focus image “objects”.	150
6.6	The arrangement of images in Figure 6.4 to 6.12.	151
6.7	Evaluation of the fusion results of night vision images with MI, Xydeas’ method, Q metrics.	161
6.8	Evaluation of the fusion results of night vision images with proposed metrics (P_{blind} , F_{blind} , and P'_{blind}).	162
B.1	The comparison of the predicted and experimental results.	183

List of Figures

1.1	The electromagnetic spectrum [2].	2
1.2	The procedure for multi-modal image fusion.	5
1.3	Two aspects of the image fusion problem.	8
1.4	The image fusion schemes.	10
1.5	The BMW night vision system on a vehicle (courtesy of BMW).	12
1.6	The organization of the thesis.	16
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.	28
2.2	The thresholded 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.	30
2.3	The top head points in two video sequences.	31
2.4	The regions of interest from two frames.	32
2.5	The refined registration results based on maximum MI.	36
2.6	The distribution of the refined scaling parameter.	37
2.7	The distribution of the refined translating parameter D_x	38
2.8	The distribution of the refined translating parameter D_y	39

2.9	The registration results. Top: IR frames; 2 nd row: EO frames; 3 th row: transformed EO frames; bottom: the synthesized images.	40
3.1	The procedure of MRA-based pixel level fusion.	46
3.2	Two images are used for testing MRA-based image fusion.	49
3.3	The fusion results with different MRA-based fusion algorithms.	50
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).	51
3.5	The visual image and infrared image.	53
3.6	The histogram-based processing of visual image.	54
3.7	The adaptive image enhancement algorithms.	57
3.8	The enhancement results of visual image achieved by adaptive algorithms.	59
3.9	The architecture of the steerable pyramid.	60
3.10	An example of steerable pyramid decomposition.	62
3.11	The pixel-level fusion of visual and IR images.	63
3.12	The result achieved by modified fusion method.	65
3.13	The enhancement function.	66
3.14	TNO Kayak (frame 7118a).	67
3.15	TNO Kayak (frame 7436a).	68
3.16	TNO Dune (frame 7404).	69
3.17	TNO Kayak (frame e518a).	70
3.18	Octec (frame 2).	71
3.19	Octec (frame 21).	72
3.20	Bristol Queen's road.	73
3.21	TNO trees (frame 4906).	74
3.22	TNO trees (frame 4917).	75

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. . . .	79
4.2	The signal processing procedures for CWD.	81
4.3	An example of image pair for CWD.	84
4.4	The image processing architectures for CWD applications.. . . .	85
4.5	The clustering indexes (a) partition index, (b) separation index, (c) Xie & Beni index, and (d) Dunn's index, with different cluster numbers.	88
4.6	The procedure for multiresolution image mosaic.	90
4.7	Illustration for accuracy and reliability assessment.	93
4.8	Multi-sensor images used for testing in the experiment: totally eight groups are involved (A-I).	94
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).	96
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.	97
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.	98
4.12	The effect of cluster number for IR image of Group A in Figure 4.8(b) . . .	99
4.13	The performance of clustering algorithms for IR image of Group A in Figure 4.8(a).	101

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.	102
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.	103
4.16	Experimental results achieved by applying the third multiresolution mosaic scheme.	104
5.1	The calculation of phase congruency map (one orientation is presented). . .	115
5.2	The P_{ref} metric for reference-based evaluation.	116
5.3	The “Gold Hill” image (left) and its phase congruency map (right).	119
5.4	The “Gold Hill” image (left) and its phase congruency map (right).	120
5.5	The “Gold Hill” image (left) and its phase congruency map (right).	121
5.6	The “Gold Hill” image (left) and its phase congruency map (right).	122
5.7	The “Lena” image (left) and its phase congruency map (right).	123
5.8	The “Lena” image (left) and its phase congruency map (right).	124
5.9	The “Lena” image (left) and its phase congruency map (right).	125
5.10	The “Lena” image (left) and its phase congruency map (right).	126
5.11	The chart for image comparison.	129
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.	131
6.1	The principal moments of phase congruency of the image in Figure 3.2(a). .	143
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.	145

6.3	The blind evaluation algorithm by using phase congruency map (P_{blind}).	146
6.4	Fusion results of image “B7118”.	152
6.5	Fusion results of image “B7436”.	153
6.6	Fusion results of image “Dune”.	154
6.7	Fusion results of image “e518a”.	155
6.8	Fusion results of image “Octec02”.	156
6.9	Fusion results of image “Octec21”.	157
6.10	Fusion results of image “Quad”.	158
6.11	Fusion results of image “Tree 4906”.	159
6.12	Fusion results of image “Tree 4917”.	160
6.13	The example of fusing a strong and a weak feature.	165
A.1	The development of phase congruency algorithm.	172
A.2	Polar diagram showing the Fourier components at a location in the signal plotted head to tail (cf. Kovesi [3]).	173
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.	175
A.4	The computation of noise compensation parameter T .	176
A.5	The 1D log Gabor filters (left: even filter; right: odd filter; top to bottom: scale from 1 to 4).	179
A.6	The even filters (left to right: scale from 1 to 4; top to bottom: orientation from 1 to 6).	180
A.7	The odd filters (left to right: scale from 1 to 4; top to bottom: orientation from 1 to 6).	181
B.1	The solid curves are obtained by five-parameter logistic regression.	184

Chapter 1

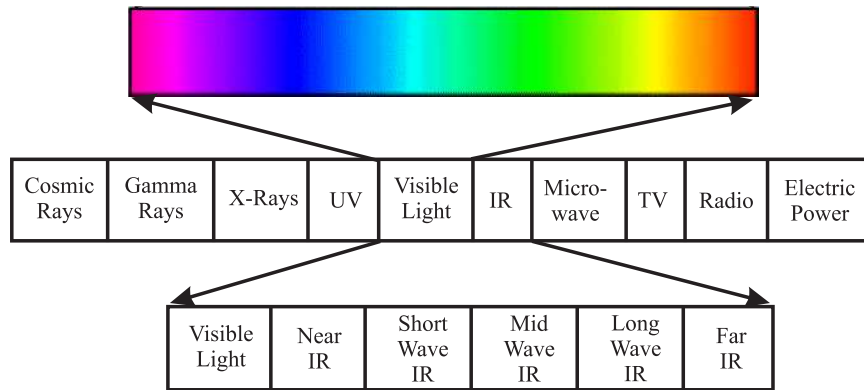
Introduction

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]:

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

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

λ : wavelength (meter)

T : Temperature (kelvin)

h : Planck's constant (joule/hertz)

c : speed of light (meter/second)

k : Boltzmann's constant (joule/kelvin)

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 pre-processing 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 world¹. Then, the registered images are combined with the fusion algorithms, which can be implemented at three different levels, i.e. pixel

¹We 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?”.

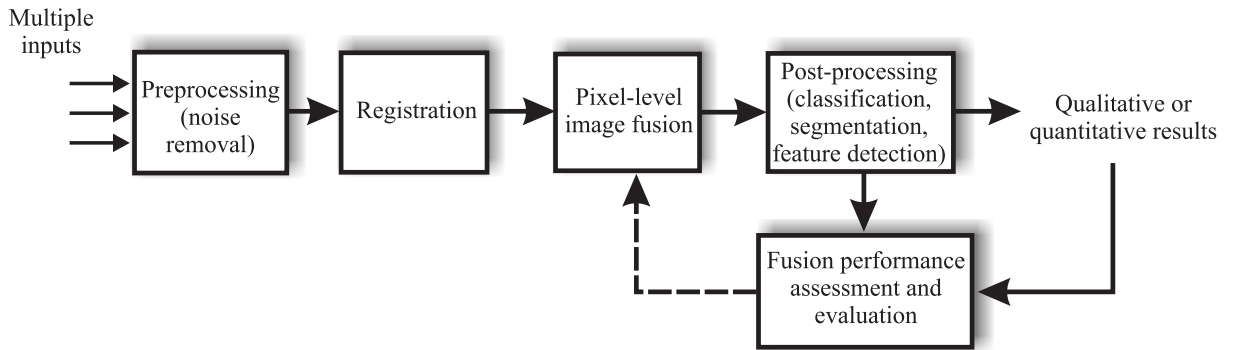


Figure 1.2: The procedure for multi-modal image fusion.

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