SUPER RESOLVED MOSAICING IN FORWARD LOOKING INFRARED IMAGERY

Thesis submitted in accordance with the requirements of the
University of Liverpool for the degree of Doctor in Philosophy

JING WANG

FEBRURARY, 2012

ABSTRACT

Forward Looking Infrared (FLIR) systems are commonly used in military applications for the purposes of detecting and recognising moving or stationary objects. FLIR imagery is particularly effective in low visibility environments and can provide additional information which would not be available in visible band images. The disadvantages of FLIR imagery are that it tends to be extremely noisy, low contrast, and cluttered due to manufacturing limitations and environmental constraints. Contemporary research has mainly focused on applying detection and recognition techniques directly to FLIR image sequences. However, compared with visible band images, FLIR imagery has much poorer quality which results in greater difficulty in detecting and recognising objects. This thesis describes the development of techniques to improve the quality of FLIR imagery prior to performing detection and recognition, with the aim of improving object detection and recognition performance. Super resolution and image mosaicing techniques have been employed for high-resolution assessment of individual areas and high-level situational awareness of large areas respectively. Both super resolution and image mosaicing rely heavily on accurate image registration hence an image registration system with sub-pixel accuracy has been developed especially for FLIR imagery. This image registration technique aligns imagery efficiently and accurately in spite of the inherent limitations of FLIR images. Then, a robust and efficient super resolution method has been adopted to enhance the image resolution and a mosaicing method based on the super resolution method used to enlarge the field of view of the image. In addition, cloud effects have been considered and a segmentation scheme developed to deal with cloud cover on FLIR imagery.

Acknowledgments

I would never have been able to finish my dissertation without the guidance of my supervisors, help from friends, and support from my family.

I would like to express my deepest gratitude to my supervisor, Dr Jason Ralph, for his full support, advice, patience and encouragement throughout the duration of my studies. He has done everything in his power for the sustainment and success of this work, which otherwise would never have been completed. I also would like to thank my secondary supervisor Dr Yannis Goulermas who was always there whenever I needed help. Many thanks also go to the other colleagues and friends who I met during my time at the University of Liverpool, especially Dr Vivek Govinda and Dr Neil Oxtoby.

Finally I would like to thank my parents and my partner for their invaluable understanding and support in numerous ways over the years.

CONTENTS

CHAPTER 1 INTRODUCTION 1

1.1 SCOPE AND MOTIVATION 2
1.2 OBJECTIVES AND METHODOLOGY 3
1.3 MAIN CONTRIBUTIONS 4
1.4 STRUCTURE OF THE THESIS 5
1.5 PUBLICATIONS 6

CHAPTER 2 INFRARED IMAGING SYSTEMS AND SIMULATION 7

2.1 INTRODUCTION 8
2.2 INFRARED IMAGING 9
2.2.1 Infrared Bands *9*
2.2.2 Infrared Imaging Systems *12*
2.2.2.1 Infrared Optics 13
2.2.2.2 Infrared Sensors 15
2.2.2.3 Read-out Circuits 16
2.2.3 Noise in Infrared Imaging Systems *18*
2.2.4 Forward Looking Infrared (FLIR) Images *20*
2.3 SIMULATED IMAGERY 21
2.3.1 Infrared Imaging System Simulations *21*
2.3.3 Simulated Target Generation *25*
2.3.4 Synthetic FLIR Imagery with Clouds *27*
2.4 SUMMARY 29

CHAPTER 3 IMAGE PRE-PROCESSING AND REGISTRATION 30

3.1 INTRODUCTION 31
3.2 CONTRAST ENHANCEMENT 31
3.3 DEAD/SATURATED PIXEL REMOVAL 36
3.3.1 Synthetic Dead/Saturated Image Processing *37*
3.4 DETECTION OF PRE-PROCESSING 38
3.4.1 Pixel Detection on Synthetic Imagery *40*
3.4.2 Pixel Detection on Real Infrared Imagery *43*
3.5 OVERVIEW OF IMAGE REGISTRATION 46
3.5.1 Differences between Images *48*
3.5.2 Image Registration Processes *49*
3.5.3 Transformation Models *51*
3.5.3.1 Rigid Transformation 52
3.5.3.2 Affine Transformation 53
3.5.3.3 Projective Transformation 54
3.5.4 Brightness Interpolation *54*
3.5.4.1 Nearest Neighbour Interpolation 55

3.5.4.2 Bilinear Interpolation 56
3.5.4.3 Bi-Cubic Interpolation 58
3.5.4.4 Other Methods of Interpolation 58
3.6 REVIEWS ON IMAGE REGISTRATION TECHNIQUES 59
3.6.1 Area Based Methods *60*
3.6.1.1 Correlation-based Methods 60
3.6.1.2 Fourier based Methods 63
3.6.1.3 Mutual Information based Methods 68
3.6.2 Feature based Methods *70*
3.7 SUMMARY 72

CHAPTER 4 ACCURATE REGISTRATION AND ROBUST SUPER RESOLUTION RECONSTRUCTION 74

4.1 INTRODUCTION 75
4.2 ACCURATE IMAGE REGISTRATION ON FLIR IMAGERY 76
4.2.1 Gross Estimation Method *80*
4.2.1.1 Pre-filtering 80
4.2.1.2 Fourier Method 81
4.2.1.2.1 Translation Registration 82
4.2.1.2.2 Rotation Registration 86
4.2.1.2 Global Image Registration 89
4.2.2 Sub-Pixel Estimation Methods *97*
4.2.2.1 Normalised Cross Correlation Method 97
4.2.2.2 Stochastic Search Method 98
4.3 SUPER RESOLUTION TECHNIQUES 101
4.3.1 Basic Theory *101*
4.3.1.1 Mathematical Formula for Super Resolution 102
4.3.2 Robust Super Resolution Reconstruction on FLIR Images *106*
4.3.2.1 Robust Super Resolution Reconstruction 106
4.3.2.2 SR Performance 110
4.4 SUMMARY 115

CHAPTER 5 CLOUD EFFECTS AND SUPER RESOLVED MOSAICING ON FLIR 117

5.1 INTRODUCTION 118
5.2 SUPER RESOLUTION TECHNIQUE ON FLIR IMAGERY WITH CLOUD 119
5.2.1 The Effect of Clouds on FLIR Imagery *119*
5.2.2 Cloud Segmentation *120*
5.2.3 Super Resolution Reconstruction on FLIR Imagery with Clouds *124*
5.2.3.1 Enhancement 125
5.3 SUPER RESOLVED MOSAICING METHOD ON FLIR IMAGERY 130
5.3.1 Image Mosaicing Theory *130*
5.3.2 Super Resolved Mosaicing Method *133*
5.3.3 Super Resolved Mosaicing Experiments *133*
5.3.3.1 Super Resolved Mosaicing on Images without Cloud 134
5.3.3.2 Super Resolved Mosaicing on Images with Cloud 136

5.4 SUMMARY 138

CHAPTER 6 REVIEW OF OBJECT DETECTION 139

6.1 INTRODUCTION 140
6.2 REVIEWS OF DETECTION ALGORITHMS 140
6.2.1 Image Thresholding 141
6.2.2 Image Differencing 142
6.2.3 Background Modelling 144

CHAPTER 7 CONCLUSIONS AND FUTURE WORK 148

7.1 THESIS SUMMARY 149
7.2 DIRECTIONS FOR FUTURE WORK 150

REFERENCES 152

LIST OF FIGURES

Figure 2.1 Division of the whole electromagnetic spectrum (Ibarra-Castanedo, 2005).....9

Figure 2.2 Blackbody emissive power spectrum (Pan and Atungulu, 2011).....10

Figure 2.3 Illustration of a CCD camera system.....12

Figure 2.4 High resolution satellite images used for generating simulated FLIR imagery: (a) suburban image (b) urban image 22

Figure 2.5 Examples of simulated infrared images (256×256): (a) with urban background and (b) with rural background.....23

Figure 2.6 Examples of simulated FLIR images: (a) simulated infrared image with noise and (b) simulated infrared image after median filtering.....24

Figure 2.7 An example of tri-bar target (Leachtenauer and Driggers, 2001).....25

Figure 2.8 Synthetic infrared image with a three bar object.....26

Figure 2.9 Examples of simulated infrared images: (a) scene without cloud cover, (b) the same scene with cloud cover.....27

Figure 3.1 Percentages under the normal curve ((Taylor, 1997).....33

Figure 3.2 Example of contrast enhancement on a 256x256 infrared image: (a) infrared image prior to contrast enhancement; (b) infrared image after contrast enhancement.....34

Figure 3.3 Histogram of the real FLIR images before enhancement and after enhancement: (a) histogram prior to contrast enhancement; (b) histogram after contrast enhancement.....35

Figure 3.4 An example of illustrating a 3×3 median filter.....37

Figure 3.5 Examples of synthetic infrared images: (a) Infrared image with dead/saturated pixels, (b) infrared image after processing by a median filter.....38

Figure 3.6 (a) An example of synthetic infrared frame without pre-processing, (b) pixel detection result with 7×7 median filter. (c) pixel detection result with 9×9 median filter, (d) pixel detection result with 11×11 median filter.....40

Figure 3.7 (a) An example of a synthetic infrared frame with pre-processing, (b) pixel detection result with 7×7 median filter, (c) detected dead/saturated pixel on dummy image (d) pixel detection result with 9×9 median filter, (e) pixel detection result with 11×11 median filter.....42

Figure 3.8 (a) An example of real infrared frame without pre-processing, (b) pixel detection result with 7×7 median filter, (c) pixel detection result with 9×9 median filter, (d) pixel detection result with 11×11 median filter.....44

Figure 3.9 (a) An example of real infrared frame with pre-processing, (b) pixel detection result with 7×7 median filter, (c) enlarged pixel detection result with 7×7 median filter, (d) detected dead/saturated pixel on dummy image, (e) pixel detection result with 9×9 median filter, (f) pixel detection result with 11×11 median filter.....45

Figure 3.10 An example of transformations between input image and template images rotation θ and translations t_x and t_y 47

Figure 3.11 An illustration of the whole region transformation.....50

Figure 3.12 Basic 2D planar transformations.....52

Figure 3.13 Nearest neighbourhood interpolation. (The discrete raster of the original image is depicted by the solid line, dashed lines show the how the inverse planar transformation maps the raster of the output image onto the input image) (Milan et al., 2008).....55

Figure 3.14 Nearest neighbour interpolation causing stepping at straight line boundaries (Zitova and Flusser, 2003).....56

Figure 3.15 Bilinear interpolation. (The discrete raster of the original image is depicted by the solid line and the dashed lines show the how the inverse planar transformation maps the raster of the output image into the input image) (Milan et al., 2008).....56

Figure 3.16 Bilinear interpolation with reduced stepping at straight line boundaries compared to nearest neighbour interpolation (Zitova and Flusser, 2003).....57

Figure 3.17 Bi-cubic interpolation (Zitova and Flusser, 2003).....58

Figure 3.18 Cross correlation search and window areas (Anuta, 1969).....60

Figure 4.1 An overview of the super resolution image reconstruction algorithm.....76
Figure 4.2 Examples of synthetic FLIR image sequences: (a) Kandarhar image sequences without cloud occlusion, (b) Kandarhar image sequences with synthetic cloud occlusion.....79
Figure 4.3 Flowchart of local Fourier registration process for translations.....84
Figure 4.4 Example of shift registration images: (a) synthetic image registration without cloud (b) registered image overlapped with the reference image.....85
Figure 4.5 Image shift registration on image sequences with cloud cover.....86
Figure 4.6 Examples of rotation registration between frames: (a) image sequences without cloud, (b) image sequences with cloud.....88
Figure 4.7 An illustration of global registration.....90
Figure 4.8 Flowchart of global Fourier registration process.....91
Figure 4.9 An example of image shift registration error on image sequences with cloud cover by aligning to the same template image.....92
Figure 4.10 Registered shifts between each frame and first frame by Fourier registration method: (a) result for image sequences without clouds by the same template Fourier registration, (b) result for image sequences without clouds by global registration, (c) result for image sequences with clouds by the same template Fourier registration, (d) result for image sequences with clouds by global registration.....93
Figure 4.11 Average error of Fourier registration on Kandarhar images.....94
Figure 4.12 An example of image registration on FLIR images with rotation and translation.....95
Figure 4.13 An example of mage registration with rotation and translations in presence of clouds.....96
Figure 4.14 An flow chart of sub-pixel estimation method.....97
Figure 4.15 Examples of cross correlation registration: (a) registration between frames without cloud (b) registration between frames with cloud.....100
Figure 4.16 Super-resolution reconstructions from a low resolution image sequence (Borman and Stevenson, 1998).....101
Figure 4.17 Observation model relating LR images to HR images (Park et al.,

2003).....103
Figure 4.18: An illustration of super resolution reconstruction.....105
Figure 4.19 An overview of flowchart super resolution reconstruction.....109
Figure 4.20 An example of the super resolution algorithm dealing with noisy images (a) the upsized noisy low resolution image and (b) the super-resolved image by using noisy low resolution images.....110
Figure 4.21 Example of (a) One of the low resolution frames and (b) High resolution image formed by 25 low resolution images.....111
Figure 4.22 Four sets of different imagery with super-resolved results. (a = Pro), (d = Kan), (g = Jal), (j = Tri) are original simulated high resolution images, (b),(e), (h), (k) are the corresponding low resolution images, (c), (f), (i), (l) are corresponding the super-resolved images.....112
Figure 5.1 A frame of FLIR imagery with cloud.....119
Figure 5.2 Optical flow image.....121
Figure 5.3 Examples of a segmented image for cases with cloud cover: (a) original simulated FLIR image without cloud, (b) original simulated FLIR image with cloud, (c) highlighted parts are the semi-transparent cloud areas, (d) segmented semi-transparent areas, (e) segmented fully clouded parts, and (f) segregated transparent/semi-transparent part of image.....123
Figure 5.4 Block diagram of proposed SR reconstruction algorithm in presence of clouds.....124
Figure 5.5 Histogram of semi-transparent areas (a) histogram for original clear cloud part, (b) histogram for original semi-transparent areas, (c) histogram for enhanced semi-transparent areas.....126
Figure 5.6 Contrast stretching result (a) original semi-transparent areas, (b) enhanced semi-transparent areas, (c) the cloud part with enhanced semi-transparent areas....127
Figure 5.7 Super resolution reconstruction result: (a) the target area in original image before super resolution, (b) the target area after super resolution.....129
Figure 5.8 An example of creating a mosaicing image from seven overlapping images.....130

Figure 5.9 (a) Super resolved mosaicing result of 20 low resolution FLIR images without cloud (without rotation), (b) Super resolved mosaicing result of 20 LR FLIR images without cloud (with rotate angle 2 degree for each frames).....135

Figure 5.10 (a)An example of the frames in the low resolution image sequences with cloud, (b)Four frames in the image sequence for the mosaicing results (Rotate angle 2 deg).....137

Figure 7.1 A system for improving the performance of detection algorithms.....150

LIST OF TABLES

Table 4.1 The correlation values of low resolution images with sub-pixel shift.....114

Table 4.2 The PSNR and correlation values for robust super resolution algorithm.114

Table 4.3 The increase values of PSNR and correlation value increase.....115

Chapter 1 Introduction

1.1 Scope and Motivation

Object detection is a technology in computer vision and image processing concerned with detecting instances of objects of a certain class in images and videos. It has wide applications in various fields. Most notably, this technology is widely used in the military to help locate and recognise targets. The challenges posed by these applications have led to a large volume of research contributions in this field, which have been mainly focussed on processing visible band imagery.

Forward-looking Infrared (FLIR) imaging is one of the most popular technologies to obtain images in dark and low visibility environments. In military surveillance, FLIR imaging systems can normally be mounted on airborne platforms to acquire ground information. However, due to limitations in the manufacturing process, infrared (IR) sensors/cameras produce images with a low signal-to-noise ratio (SNR) and their sensor arrays have inconsistent sensitivities from pixel to pixel. In addition, as the FLIR imaging is based on the objects` temperature, it is very low contrast and easily cluttered.

Compared with visible band images, FLIR images have much poorer quality which makes detection tasks harder to accomplish. Instabilities are also introduced when FLIR devices are mounted on mobile platforms such as aircraft, which also adds the problem that objects on the ground appear relatively small, whose size may only cover a few pixels and as a result different ground features and coverings can add difficulties to the detection task.

1.2 Objectives and Methodology

Contemporary research has mainly focused on applying detection and recognition techniques directly to FLIR image sequences. However, compared with visible band images, FLIR imagery has much poorer quality which results in greater difficulty in detecting and recognising objects.

The objective of this thesis is to develop techniques to improve the quality of FLIR imagery prior to performing detection and recognition in order to improve object detection and recognition performance. As both super resolution and image mosaicing rely heavily on accurate image registration to align images, an image registration system with sub-pixel accuracy needed to be developed to be used on FLIR imagery. Following this, a robust and efficient super resolution method had to be adopted to enhance image resolution and a mosaicing method based on the super resolution method to enlarge the field of view. Super resolution and image mosaicing techniques have been employed for high-resolution assessment of individual areas, and high-level situational awareness of large areas respectively.

The Matlab programming environment has been chosen for testing and developing the algorithms used in this thesis as it allows quick insights and fast prototyping.

1.3 Main Contributions

There are three main contributions in this thesis. First, a sub-pixel accuracy image registration system has been developed for FLIR imagery. This registration system combines the Fourier registration method and maximum cross correlation method to achieve sub-pixel accuracy. The Fourier registration method is used for large movement and is robust and efficient for FLIR imagery. The maximum cross correlation method is used to adjust residual alignment.

Second, a fast super-resolved mosaicing technique has been developed specifically for FLIR imagery to enhance image resolution and expand the field of view of the image.

Thirdly, a segmentation method has been developed especially for use in imagery with cloud cover. This method can segment the transparent and semi-transparent parts of the image, which can then be further processed, whilst discarding the parts of the image that cannot be used.

1.4 Structure of the Thesis

This thesis is composed of the following chapters:

- Chapter 2 presents background information on infrared imaging systems and describes the development of simulated infrared imagery in different environments.
- Chapter 3 describes some image processing techniques which can be used for pre-processing and consequently improving the quality of infrared imagery, so as to remove noise, improve contrast and better enable feature identification.
- Chapter 4 examines the image registration algorithms on FLIR image sequence and develops a sub-pixel registration system for FLIR images in the presence of large amounts of noise and cloud. Then super resolution algorithms are investigated and a robust super resolution method is applied to FLIR images.
- Chapter 5 describes the effect of clouds on FLIR imagery. Both situations that are with clouds and without clouds are considered. Then a super-resolved mosaicing method is employed to generate a high resolution panoramic image from a sequence of low resolution FLIR images.
- Chapter 6 gives an overview of the main issues and research regarding small object detection and serves as an introduction to the potential applications provided by this research.
- Chapter 7 provides a thorough discussion of conclusions and future work leading on from this research.

1.5 Publications

Parts of this thesis have been presented at conferences.

- J. Wang, J.F. Ralph, J.Y. Goulermas, 'An analysis of a robust super resolution algorithm for infrared imaging', Proceedings of 6th International Symposium on Image and Signal Processing and Analysis, ISPA 2009.
- J. Wang, J. F. Ralph, and J. Y. Goulermas, 'High-resolution reconstruction of objects from cloud-covered infrared images', Proc. SPIE on Signal Processing, Sensor Fusion, and Target Recognition XVIII, Vol. 7336, 2009.
- J. Wang, S. Bao, J. F. Ralph, J. Y. Goulermas, 'Detection of small objects in multi-layered infrared images', Signal and Data Processing of Small Targets 2008 Proceedings of SPIE Vol. 6969.

Chapter 2 Infrared Imaging Systems and Simulation

2.1 Introduction

Although imaging systems were first developed using only visible light, the use of the wider electromagnetic spectrum makes it possible to capture a wider variety of images. One of the most useful imaging systems uses infrared light (also known as thermal imaging), and has been widely used in many different fields. Infrared imaging systems operate on the principle that all objects emit infrared energy as a function of their temperature. An infrared imaging system collects infrared radiation from objects in the scene and then creates an electronic image, where the brightness of an object is related to its temperature. Infrared images tend to be monochromatic because wavelengths outside the normal visible spectrum do not map uniformly into the system of colour vision used by humans. Compared to imaging systems that use only visible light, the main advantage of an infrared imaging system is that it can be used in dark conditions without needing to enhance ambient visible light.

Infrared imaging systems have a broad range of applications, including navigation, night security and medical imaging, as well as military applications such as reconnaissance and surveillance. The most challenging environments and most demanding requirements tend to be found in military applications, and this is the main context of this research. The military is interested in detecting, recognising and identifying targets at long distances, an application which requires high-resolution, low-noise sensors. Modern military applications that use infrared imaging utilise various wavelengths ranging from mid-wave infrared to very long wave infrared (Li, Liu and Tidron, 1998).

This chapter gives basic background information on infrared emission and detection methodologies, before then describing the formation of synthetic Forward Looking Infrared (FLIR) imagery and the generation of targets which are used as test imagery in later chapters.

2.2 Infrared Imaging

2.2.1 Infrared Bands

Visible band imaging systems are most familiar to us since our human eyes are good imagers in the visible band of the electromagnetic spectrum. The infrared waveband extends from wavelengths of 0.74μm to 1mm while visible band is from wavelengths of 0.35μm to 74μm, as shown in Figure 2.1.

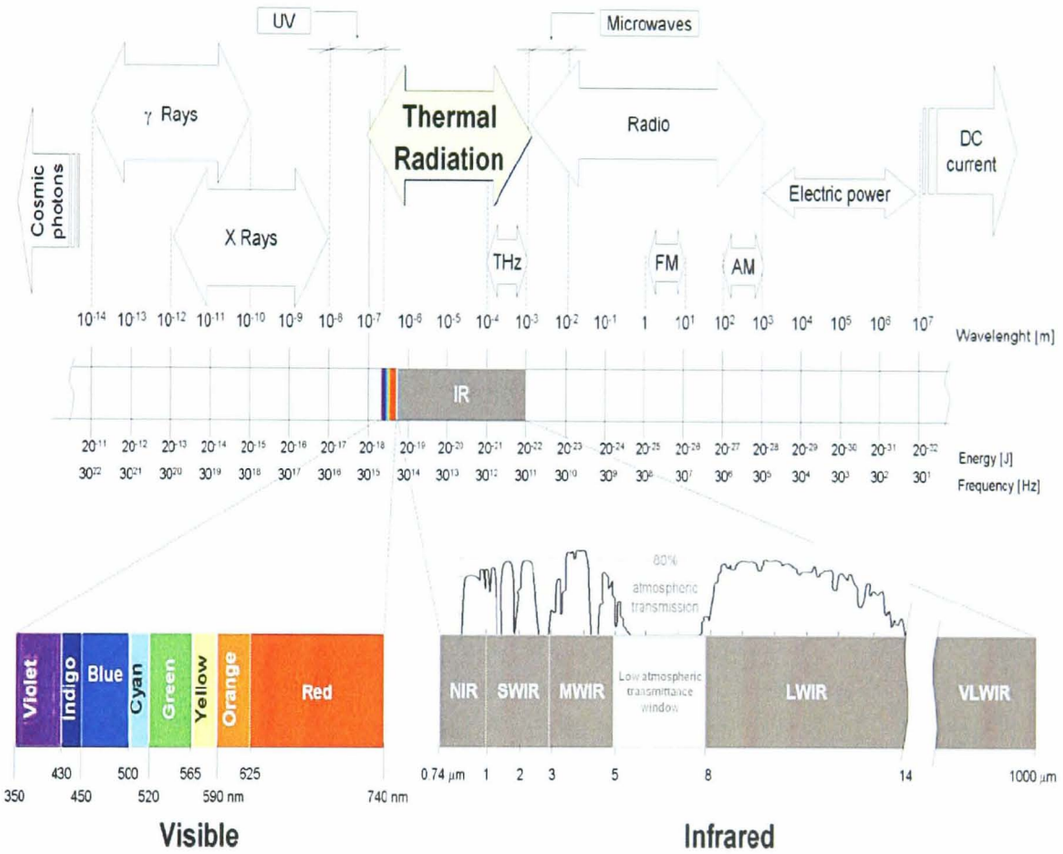


Figure 2.1 Division of the whole electromagnetic spectrum (Ibarra-Castanedo, 2005)

The infrared portion of the electromagnetic spectrum can be split into several wavebands. The near infrared (NIR) band is around 0.74 – 1.0μm; the infrared radiation of short-wavelength (SWIR) has a wavelength of up to 3.0μm; the wavelengths between 3.0 – 5.0μm are mid-wave infrared (MWIR); long-wave is 8.0 – 14μm; and wavelengths up to 1000 μm are termed very long-wave infrared (VLWIR).

As atmospheric constituents do not absorb evenly, the infrared waveband is not a uniform region. Therefore, among the infrared bands, the waveband between 5.0 – 8.0μm and a small band of 4.0 – 4.2μm are very strongly absorbed by the atmosphere. This is because when electromagnetic radiation travels through the atmosphere, it may be absorbed or scattered by constituent particles and the overall effect of this is to remove energy from the radiation. The dominant factor in atmospheric absorption is water vapour, followed by carbon dioxide and then ozone (Kirill, 1969).

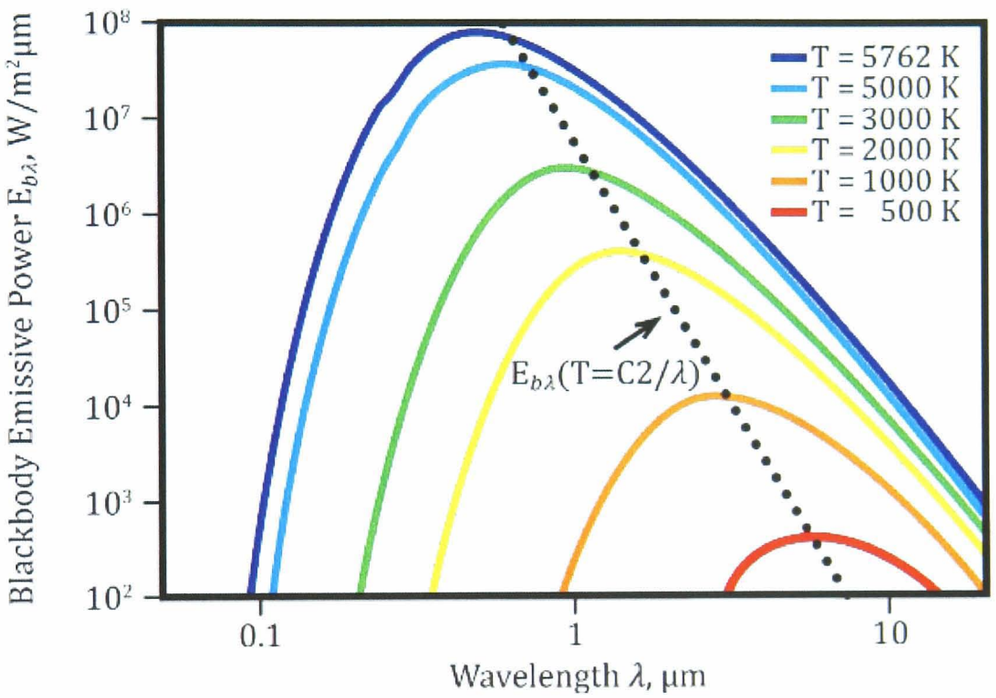


Figure 2.2 Blackbody emissive power spectrum (Pan and Atungulu, 2011)

The difference between the infrared bands and the visible band can also be assessed by the thermal emission. A black-body is an idealised physical object that absorbs all incident electromagnetic radiation that falls on it. Because of this perfect absorptivity at all wavelengths, a black-body is also the best possible emitter of thermal radiation. The temperature of a black-body determines the radiation that is emitted; the black-body radiation curve is shown in Figure 2.2. The dotted line in Figure 2.2 demonstrates the relationship between the source temperature and the peak wavelength. In reality, an object will absorb and emit certain wavelengths more or less strongly depending on the structure of the material.

Spectral radiance can be used to measure the amount of radiation such as light or radiant heat that passes through or is emitted from a particular area. At a given temperature, the spectral radiance $I(\lambda,T)$ of electromagnetic radiation at all wavelengths emitted in the normal direction from a black body can be calculated by Planck's formula:

$$I(\lambda,T)=\frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda k_B T}-1} \tag{2-1}$$

where λ is the wavelength, $h=6.626\times10^{-34}Js$ is Planck's constant, $c=2.998\times10^{8}ms^{-1}$ is the speed of light and $k_B=1.381\times10^{-23}JK^{-1}$ is Boltzmann's constant. By integrating I over the solid angle subtended by the imager optical aperture and over the projected area of the source and the wavelength interval to which the camera is sensitive, one obtains the total radiated power received by the detector before such attenuating factors as optical transmission and atmospheric losses are taken into account (Ralph and Bernhardt, 2002).

Lower temperature objects emit a negligible amount of visible light. However, using the infrared spectrum, objects too cold to emit visible light can still be seen to radiate in the infrared. Both visible and infrared radiative emission is temperature dependant; hotter objects will radiate more electromagnetic energy than cooler ones.

Thermal imaging sensors are traditionally built to operate in specific bands. The selection of wave band depends on the application, for example MWIR is mainly used for hot and humid environments, whereas for cold and dusty environments the LWIR band is preferred.

2.2.2 Infrared Imaging Systems

The basic structure of an imager consists of an optical system (lenses), photosensitive sensors and electronic read-out circuits. A block diagram of a Charge Couple Device (CCD) camera system for visible band detection is shown in Figure 2.3. The front end optical sub-system focuses incident radiation onto the focal plane of the sensor. Subsequently the sensor, an array of photo-detector elements, converts the incident radiation into an electrical signal. The AGC (Automatic Gain Control) element automatically changes the gain of the camera according to the amount of light in the scene. The gain is set as a compromise between necessary sensitivity in low illuminated areas and the attempt to avoid saturation of bright areas in the scene. The high-pass filter in the camera compensates for a decrease in high frequencies in the optics signal. The output is processed by a set of read-out electronics that process the signal and controls the gain and bias of the detectors. The signal can then be passed by these read-out circuits to an analogue-to-digital converter, which digitises the signal to a finite number of bits for further processing or storage (Sonka et al., 2008).

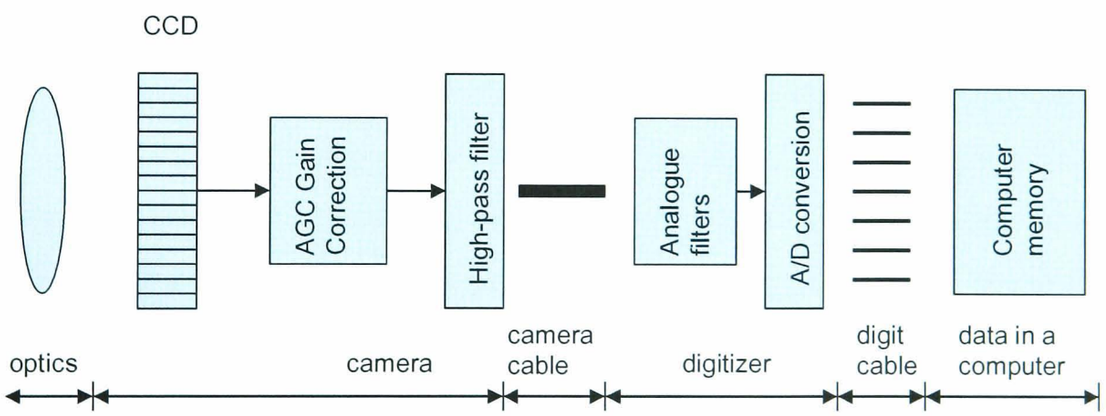


Figure 2.3 Illustration of a CCD camera system

Infrared imaging systems generally have the same structure as a visible band imager as shown in Figure 2.3. However, they have more specialised requirements in terms of the material for optics and the detector read out circuit. More detailed explanations will be given in subsequent sections.

2.2.2.1 Infrared Optics

The optical system has a direct effect on the performance of an imaging system. A well designed optical system can provide a good response which can then be converted into a better electronic response. The design of camera optical systems can be quite complex and usually involves the use of several different lenses to control chromatic aberrations, coma and distortion (Holst and Lomheim, 2007).

Chromatic aberration is a type of distortion caused by the differing refractive indices of varying wavelengths of light, refractive index decreases with increasing wavelength (Gross et al., 2007). Chromatic aberration is resolved by the use of a lens system that focuses all wavelengths to the same convergence point. Coma aberration is an optical effect caused by imperfections in the lens or other components, and causes off-axis point sources to appear distorted. In optical systems imaging a wide spectral range, coma can be a function of wavelength, in which case it is a form of chromatic aberration (Zappe, 2010).

Infrared optical systems are more expensive than those used for visible light applications due to the fact that conventional optics are not suitable for the role. Infrared imaging systems require the material of the lens to be very transmissive in the infrared band. The available optical materials for infrared systems vary according to their intended waveband application. For example, Ge (Germanium), ZnS (Zinc Sulphide), ZnSe (Zinc Selenide), AMTIR-1 (Amorphous Material Transmitting Infrared Radiation) are used for the 8.0 – 12.0μm band; while Germanium, ZnS and Si (silicon) are used for 3.0 – 5.0μm waveband (Singh, 2009).

Another consideration in infrared systems is aperture size. In most visible light applications, aperture size is not a restriction because the wavelength of visible light is relatively small with respect to the optics; however the situation is different in infrared. In infrared applications, the effect of aperture size varies between

wavelengths, which have larger effects in the near infrared region than far infrared (Ralph and Bernhardt, 2002).

The angular resolution of a camera is restricted by the size of the aperture and the available magnification of the optics. The angular resolution of a circular lens can be determined by the Rayleigh criterion:

$$\Delta\theta = 1.22\lambda / D \tag{2-2}$$

where $\Delta\theta$ is the angular resolution in radians, λ is the wavelength of light, and D is the diameter of the lens' aperture.

In the infrared waveband, the resolution of a camera is frequently limited by the size of the aperture, because it is difficult to make large optical elements from materials suitable for infrared (Mooney and Sheperd, 1996). Resolving small objects at long range requires a narrow field of view because there are only a fixed number of detector pixels on a sensor; therefore, according to the Rayleigh criterion, it is desirable to have a larger aperture.

In military applications MWIR and LWIR are the most common types of waveband detectors used (Maurer et al., 2009). The biggest difference between these systems is the size of the detector apparatus. This is because spatial image resolution is a function of pixel pitch and field of view, and longer wavelengths such as in LWIR, have a larger diffraction spot size. This is compensated for by matching the spot size to the pixel pitch, if it were not, it would lead to images appearing blurry. This means that the physical size of the LWIR pixels needs to be larger, along with the associated optics. As a consequence, MWIR is often preferred because it allows higher resolution for a given aperture size and means detectors are easier to build and cameras are smaller. Infrared images tend to suffer from low contrast due to the fact that objects in a scene tend to be of similar temperature; therefore the intensity of infrared emission is similar and it is a small difference on a large background signal.

2.2.2.2 Infrared Sensors

Visible imaging systems use the same silicon technology for sensors and read-out circuits, which enables a high degree of component integration and less noise introduced by the read-out process. By contrast, infrared imaging systems are more complex as they use different materials for photo-detectors and read-out circuits and consequently it is necessary to distinguish between the two.

Although a range of infrared detector technologies have been developed, the two of most interest in contemporary systems are semiconductor bandgap detectors and newer quantum well detectors (Li et al., 1998, Kopp, n.d.).

Bandgap detectors are made of semiconductor materials which require a photon to have a certain wavelength or energy to dislodge an electron from the crystalline lattice of the material and produce a measurable electrical effect (Kopp, n.d.). To detect infrared radiation, narrow band-gaps among materials are required. CdHgTe/Cadmium Mercury Telluride (CMT) is an example of a bandgap detector used in most early imaging systems, and is especially used in long wave applications (Tidrow and Dyer, 2001). CMT detectors have high quantum efficiency (the number of electrons generated for every incident photon), typically around 75-80%, and are fast in operation (Galileo, 2007; Smith, 1995). The disadvantages of CMT detectors are that they have a limited uniformity caused by defect related dark current, and they are also difficult to fabricate and integrate into focal plane arrays, such as CCDs (Tidrow and Dye, 2001). As a consequence, in current devices CMT has to be fabricated separately from the silicon read-out circuits and then bonded to the silicon circuits using indium 'bumps' which leads to dead/saturated circuit pixels. Indium Antimonide (InSb) is much easier to fabricate than CdHgTe, but it is limited to midwave infrared applications. Platinum Silicide (PtSi) is an example of another material used, specifically for shortwave infrared detector systems (Abraham et al., 1988). Most current applications will use one of these three materials.

The new Quantum Well Imaging Photodetectors (QWIP) are quite different from previous technology and offer some advantages over bandgap detectors. In QWIP, a well-shaped microscopic hole is created in a semiconductor material. typically a Gallium Arsenide (GaAs) alloy. If a photon of suitable energy falls into the well, electrons can be dislodged; otherwise the electron will be trapped in the well. The mass production of QWIPs means they are potentially much cheaper, especially as they are already widely used in radio-frequency chips in mobile phones, wireless networks and radar (Kopp, n.d.). In terms of disadvantages, QWIPs suffer from excess dark current which is a strong function of operating temperature. To keep the dark current low and prevent saturating the storage of each pixel during the stare time, the QWIP normally needs to be cooled to around 60-65K, depending on the peak wavelength. Although QWIPs represent the new frontier in infrared imaging systems, they are relatively inefficient compared to established bandgap detectors. The quantum efficiency is normally around 5-10% (Galileo, 2007). This means that QWIPs require longer stare times to get a useful signal in the pixel capacitor.

Both CMT and QWIP detectors have potentially desirable characteristics and capabilities for infrared detectors. CMT is more applicable where there is a requirement for very high sensitivity, relatively high operating temperatures and high speed, whereas QWIP is more suitable where very large format, high uniformity, very long wavelength, and low temperature operation are required (Li, et. al, 1998).

2.2.2.3 Read-out Circuits

In visible light imaging there are two main types of semiconductor sensor used, namely charge couple devices (CCD) and complementary metal oxide semiconductor (CMOS). Although the two types of technology were developed at around the same time, charge couple device technology matured at a faster rate and consequently became more widely used.

The basic functions of a CCD sensor are charge collection, charge transfer and the conversion of charge into a measurable voltage. After a photon is absorbed by the photo-detector, an electron-hole pair is created. In a CCD sensor, every pixel's charge is transferred through just one output node to be converted to a voltage, buffered, and sent off-chip as an analogue signal. A CCD sensor can be considered to contain a two-dimensional array of pixels linked together. All of the pixel area can be devoted to light capture (McCann and Scott, 2005, Sonka et al., 2008).

In a CMOS sensor, each pixel has its own charge-to-voltage conversion, and the sensor often includes amplifiers, noise-correction and digitisation circuits. These other functions increase design complexity and reduce the area available for light capture. However, the chip can be built to require less off-chip circuitry for basic operation (Sonka et al., 2008).

There are different advantages and disadvantages for CCD and CMOS sensors in capturing images depending on the application. Some advantages of a CCD sensor over CMOS are that it does not obscure a large photoconductive area on the chip. Second, it can be used for detector arrays with many pixels enabling it to have a larger fill factor than CMOS. Moreover, the exposure time of all pixels is uniform and is independent of readout time (McCann and Scott, 2005).

However, CMOS technology has the advantage of allowing for far more sophisticated storage, multiplexing and other manipulation including random pixel access and reading out sub-images, or combining pixels to form lower resolution images, but it requires a larger area on the silicon chip. It is not possible to allow random access to pixels in CCD systems because all of the pixels are linked together. (Ralph & Bernhardt, 2002). Further advantages of CMOS are that they can operate at very high speed and have lower power consumption and inexpensive fabrication (Siegwart & Nourbakhsh, 2004).

2.2.3 Noise in Infrared Imaging Systems

Two terms can be used to describe variation in arrays and sensors, namely non-uniformity and spatial noise (Mooney and Sheperd, 1996). Non-uniformity describes the variation in the time averaged output of the detectors in an infrared focal plane array, and spatial noise refers to the post compensation noise observed in an infrared image.

The term non-uniformity is usually used to describe the focal plane arrays. The non-uniformity of the detector array is widely recognised as the main factor in limiting infrared sensor performance (Mooney and Sheperd, 1996). As mentioned previously, in standard bandgap semiconductive system, the infrared photo-detector material has to be fabricated separately from the silicon read-out circuits and then bonded to the silicon circuits using indium 'bumps'. This process has difficulties when aligning a large number of indium bumps in the appropriate position on the read-out circuits, especially for large arrays.

Non-uniformity can be generated by poor connections between photo-detector and read-out circuits, and can be due to inconsistencies in photo detector response due to manufacturing tolerances. This non-uniformity is a major problem in nearly all infrared imaging systems; however the development of digital processing allows a wide range of non-uniformity correction techniques which can improve image quality. The usual way to deal with non-uniformity is to use a uniform blackbody source to find the response at each pixel and then scale the response to see whether there is a difference in a particular pixel (Ralph & Bernhardt, 2002). Although non-uniformity correction can dramatically enhance infrared camera performance, operational constraints and cost of a particular application often limit the degree of non-uniformity correction.

Spatial noise is caused by random variations of camera operating conditions and unexpectedly changing environments. Fixed pattern noise, also known as salt and pepper noise, is a type of spatial noise that is frequently present in infrared images.

Noise in an infrared imaging system is usually characterised in terms of a few simple parameters. The noise equivalent temperature difference (NETD) is the oldest and most widely used measure of the ability of a system to discriminate small signals in noise. The simplest and most commonly used definition follows: the NETD is the blackbody target-to-background temperature difference in a standard test pattern which produces a peak-signal to rms-noise ratio (SNR) of one at the output of a reference electronic filter when the system views the test pattern (Lloyd, 1975). It describes the noise because it combines a range of different noise sources into a single parameter. For a high-performance infrared imaging system, an NETD around 0.05 K is fairly typical (Ralph & Bernhardt, 2002). Other terms used to characterise the degree of non-uniformity and spatial noise are fixed pattern noise, synthetic pattern noise, residual non-uniformity, residual spatial noise and uniformity.

There are a large number of noise sources in infrared imaging systems and they appear in nearly every component of the system. The sources of noise can be divided into scintillation effects, due to the discrete nature of the electrons; optical distortion, detector and circuit non-uniformities; and quantisation effects from the digitisation of the output. Although the origins of the noise sources are different, they all appear as variations in the image intensity and prevent the formulation of a clear image (Holst & Lomheim, 2007).

2.2.4 Forward Looking Infrared (FLIR) Images

Forward Looking Infrared (FLIR) imaging systems are one of the most widely used high quality infrared application used at present. FLIR systems are often used on satellites, naval vessels, fixed-wing aircraft, helicopters and armoured fighting vehicles since they can detect objects through smoke, fog, haze and other atmospheric obscurants better than visible light cameras. The term "forward looking" was initially used to distinguish FLIR systems from sideways tracking infrared systems. However, nowadays the term is used for most kinds of military infrared imaging system.

A complication associated with FLIR infrared systems is that they tend to be mounted on a moving platform; consequently there are aerodynamic issues which normally require the constituent optical components be small and robust. Moreover, under difficult circumstances FLIR imagery may have abrupt discontinuities in motion. There is another problem tend to occur in FLIR images, the object edges and corners of the images tend to be smoothed out, leading to a reduction in distinct features due to the filter built in the camera.

Moving objects are often much brighter than the background as a result of environmental heating effects such as friction with air and engine combustion. (Sterel and Aggarwal, 1999). However, there are also large amount of noise among some of them are bright saturated noise which can make the images cluttered. When the size of object is small, the noise can cause false alarm in the object detection process.

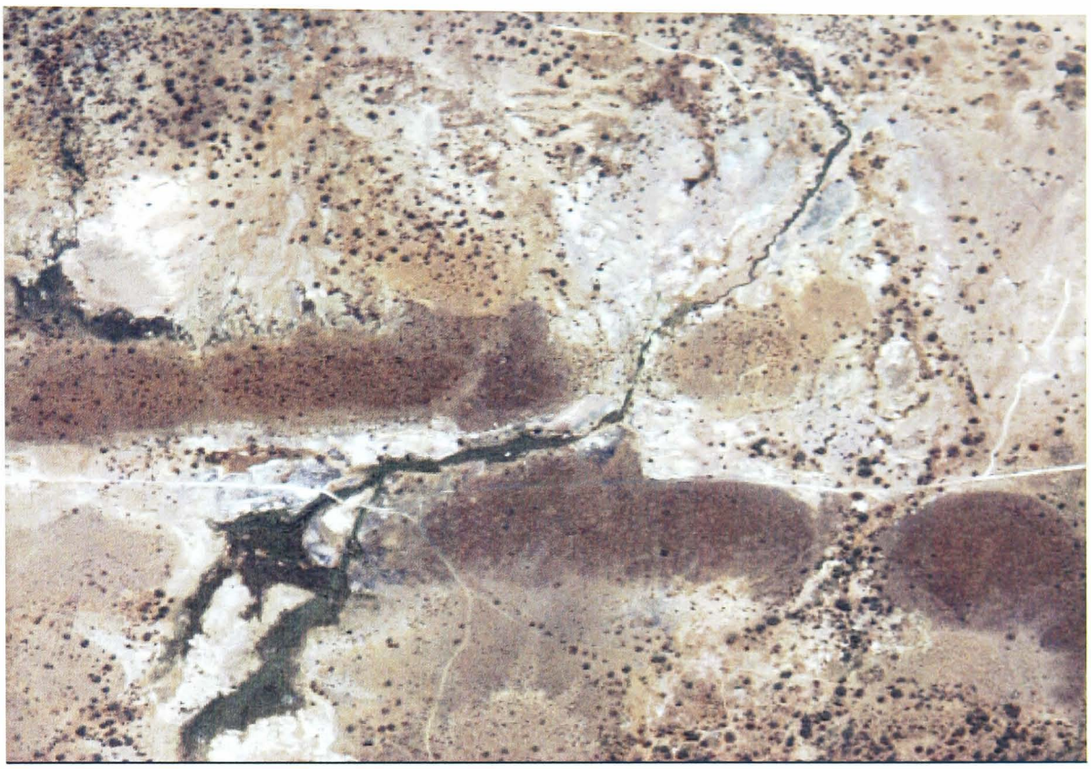
Due to the inherent limitations of infrared images, as well as the added operational aspects of FLIR systems, improving this type of imagery using signal processing is very challenging.

2.3 Simulated Imagery

2.3.1 Infrared Imaging System Simulations

Simulated infrared imagery was formed so as to form a basis upon which the algorithms developed in this thesis could be applied and tested with the aim of improving the quality of the FLIR images. A basic simulated infrared imaging system model was utilised which was based on an existing physics-based camera model, more detail of the model can be found in the paper by Ralph et al. (2005). It used high resolution satellite imagery to initially create a ground scene then, so as to provide the correct statistical variation; the images were modified to include the standard types of noise and distortion present in real infrared cameras, such as dead/saturated pixels, non-uniform pixel responses and automatic gain-offset correction.

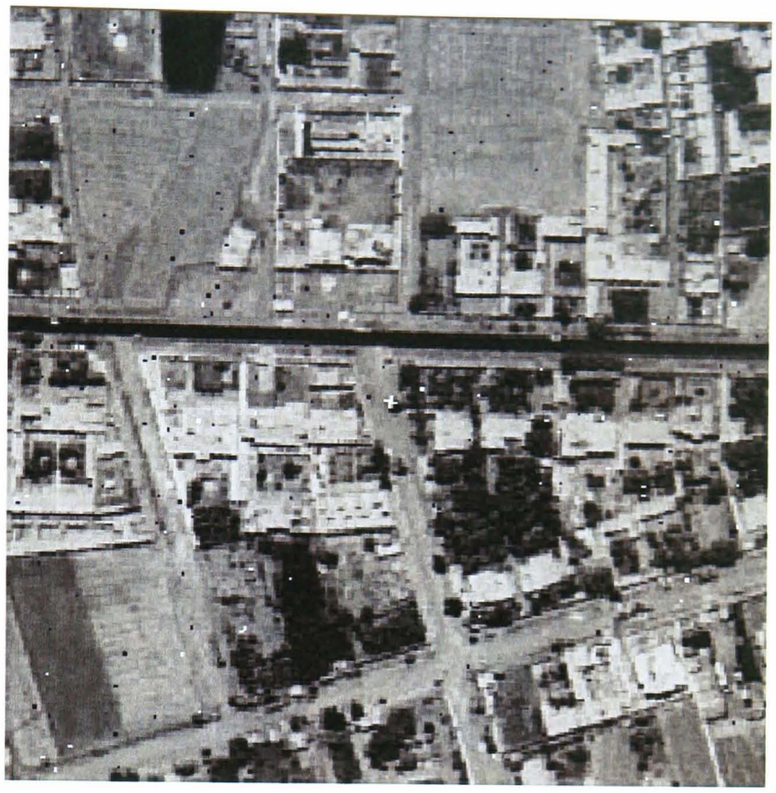
Whilst this camera model is not fully realistic, it does embody most of the limitations of real infrared cameras and provides a source of image data allowing PC-based closed-loop simulations of aircraft camera systems. Infrared image sequences were generated by the simulated camera mounted on a simulated moving and/or vibrating platform representing an aircraft mounted camera. The standard camera was taken to be a downward-looking, mid-wave IR camera with a 256×256 pixel focal plane array, a 16 degree field of view, 25-50 Hz frame rate and a Noise Equivalent Temperature Difference (NETD) of 0.5K. The speed and direction of the moving platform could be controlled to generate imagery for different requirements. However, camera parameters could be varied as required. Figure 2.4 shows two background images with a suburban and urban scene respectively. They are used for generating simulated FLIR imagery. Figure 2.5 shows the simulated infrared images after processing by the simulation algorithms. The simulated infrared images are contaminated by a large amount of noise, representing real infrared images.



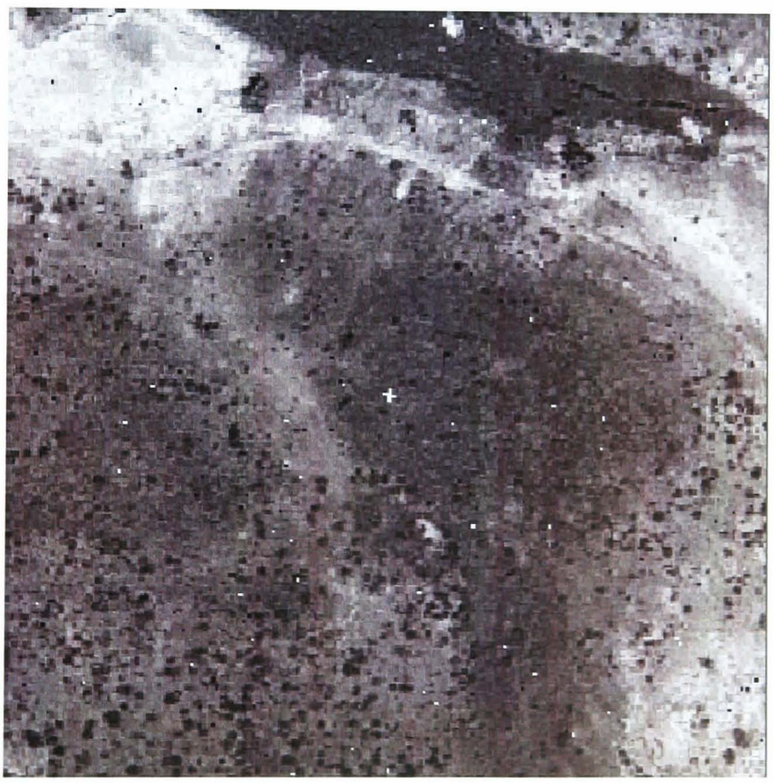
(a)

(b)

Figure 2.4 High resolution satellite images used for generating simulated FLIR imagery: (a) suburban image (b) urban image (http://uuu.mindtel.com/2006/0307.jalalabad/maps/)

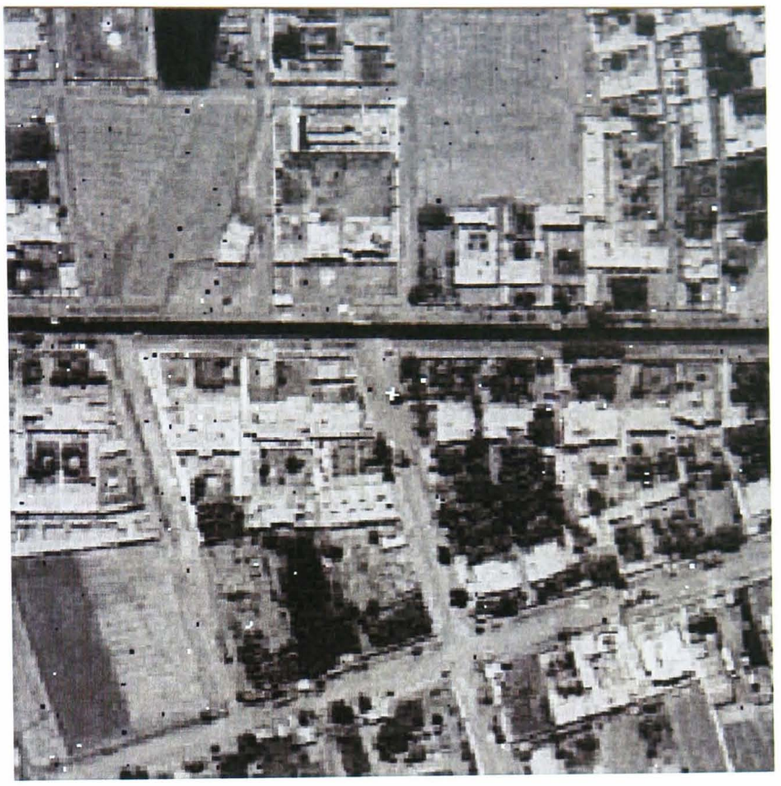


(a)

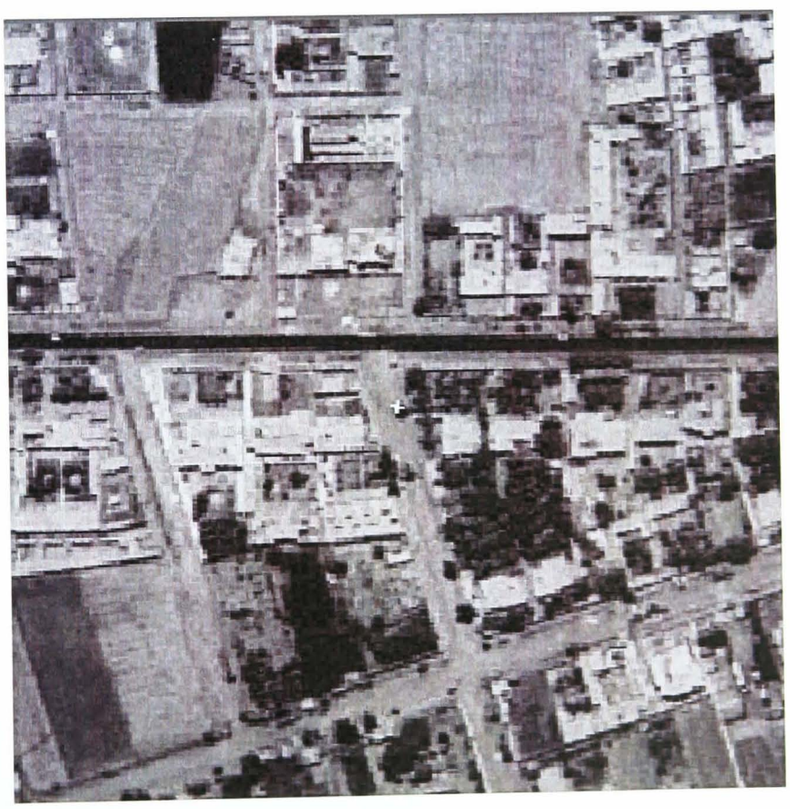


(b)

Figure 2.5 Examples of simulated infrared images (256×256): (a) with urban background and (b) with rural background.



(a)



(b)

Figure 2.6 Examples of simulated FLIR images: (a) simulated infrared image with noise and (b) simulated infrared image after median filtering

This camera model also contains a built-in function for a median filter, which removes dead/saturated pixels from the infrared image. Normally, this median filter is only applied to the dead or saturated pixels otherwise the whole image would be blurred. The effect of the median filter can be observed in Figure 2.6.

2.3.3 Simulated Target Generation

A target area is typically identified as a region of the image brighter than the background since in military applications the targets are generally hot objects. In the simulation, the size and initial position of the target can be set in the program and its movement also can be set separately from the camera movement by its own motion algorithms.

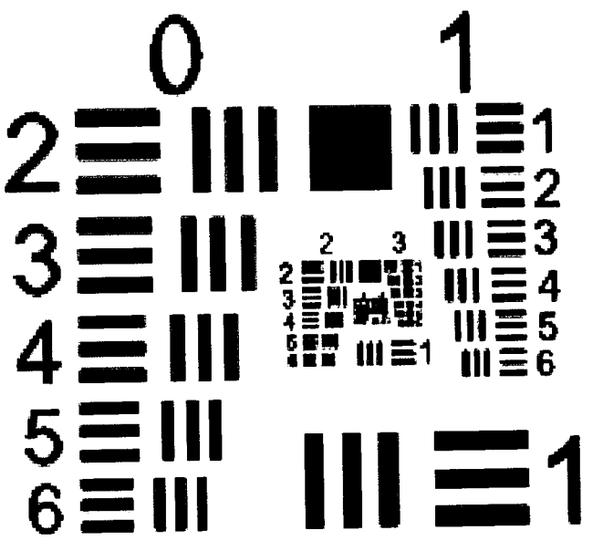


Figure 2.7 An example of tri-bar target (Leachtenauer and Driggers, 2001).

Target acquisition is generally concerned with the detection of points of interest (POIs) and subsequent recognition and identification. The ability to detect and recognise an object in monochrome can be set by the Johnson criterion. These criteria sets up the relationship between the ability of human observers to resolve bar targets (one black bar and one white bar equate to one cycle) through an imaging device and their ability to perform the tasks of detection, recognition and

identification of military vehicles through the same optical sensor (Howe, 1993). Figure 2.7 shows an example of a tri-bar resolution target (Leachtenauer and Driggers, 2001). The Johnson criteria are described as follows:

1. Detection (with 50% probability) = observer able to resolve 1 bar

2. Recognition (with 50% probability) = observer able to resolve 3 bars

3. Identification (with 50% probability) = observer able to resolve 5 bars

The simulated small bright object is combined in the background scene as seen in Figure 2.8. The object is a square three bar target with a bar length of a = 5-10 metres and a bar width of $a/5$ and in some cases an additional bar is added to ensure that the target bars are linked as seen in Figure 2.8. When the simulated air platform moves at a certain speed, the attitude of camera model changes at a certain or random degree, with the bright object moving at a certain speed and direction. Consequently, the generated imagery can be considered to represent a moving camera tracking a moving object.

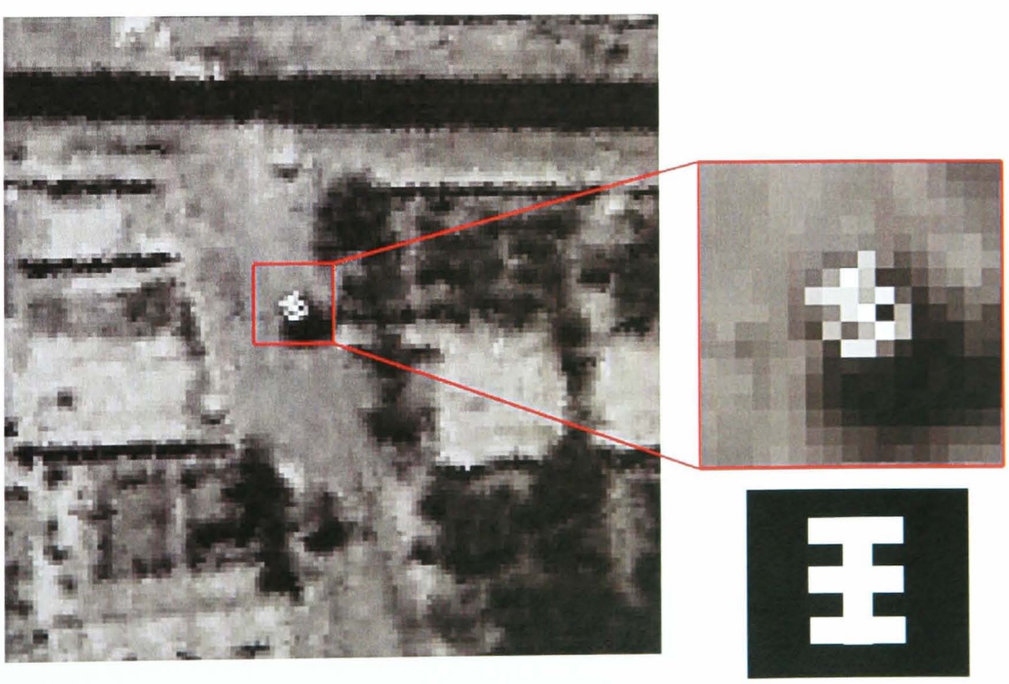


Figure 2.8 Synthetic infrared image with a three bar object

2.3.4 Synthetic FLIR Imagery with Clouds

As well as being low contrast and noisy, another challenge of FLIR imagery is due to cloud occlusions. Clouds tend to obscure ground features and reduce contrast on regions that are cloud-free. In order to simulate this situation, a cloud model has been applied to generate clouds on the synthetic FLIR imagery. The cloud model is generated based on a statistical infrared clutter model proposed by Itakura et. al (1974). Since the colour of the cloud in the image is determined by the cloud temperature, it can appear brighter or darker than the background image. When the temperature of cloud is higher than the ground, it appears brighter in the image. When the temperature of cloud is lower than the ground, it appears darker. However, the shape and the thickness of the cloud will remain the same. This cloud model generates white cloud with random shape and thickness. It provides a relatively realistic simulation of the clouds.

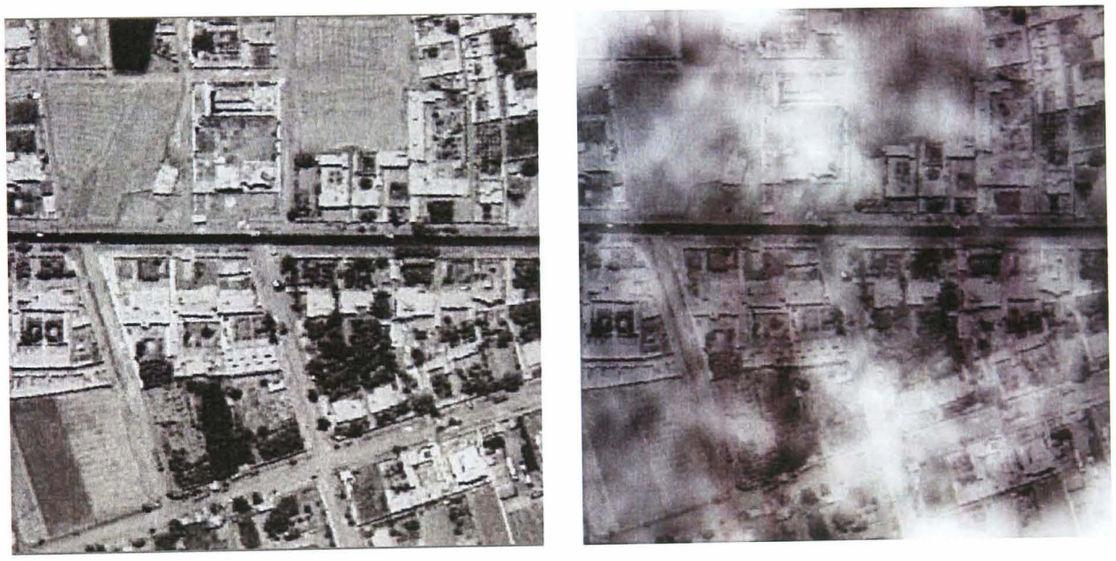
Clouds are generated based on a random set of the two-dimensional pulses whose amplitude and width follow Gaussian and Poisson statistics, respectively:

$$p(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(n-\bar{n})^2}{2\sigma^2} \right] \tag{2-3}$$

$$p(r) = \alpha \exp(-\alpha r) \tag{2-4}$$

where n is the radiance of a certain point on the x-y plane, $\bar{n}$ is the mean value of n, σ^2 is the variance of n, r is the interval length between two adjacent points on the x-y plane, and α is the reciprocal of the average pulse width.

The cloud models are then inserted into the previously discussed camera model as a three-dimensional structure at a defined height above the ground plane. By varying the height of the clouds relative to the ground and the aircraft/imager, the effect of parallax on the ego-motion can be assessed. The statistical properties specified in the model can be used to control the amount and density of the cloud cover. The parameters have been selected to represent about 2-4 Oktas of stratus clouds – more cloud cover than this makes it very difficult to track an object on the ground for long enough to use super-resolution techniques. The images produced in this way are relatively realistic, as shown in Figure 2.9 (b) and embody many of the infrared properties of real clouds.



(a) (b)

Figure 2.9 Examples of simulated infrared images: (a) scene without cloud cover, (b) the same scene with cloud cover.

2.4 Summary

The use of the infrared portion of the electromagnetic spectrum is advantageous in certain applications because it allows to display features in the environment either not easily observed using visible light systems or totally dark. Despite this, there are certain limitations related to infrared imaging technology, such as relative low resolution and much noisier than visible band image. Infrared systems also tend to be more complex and expensive due to material constraints in building systems capable of detecting infrared scenes. Image processing techniques can be used to post process and consequently improve the quality of infrared imagery, so as to remove noise, improve contrast and better enable feature identification.

In this research, a means of generating synthetic infrared imagery from a simulated aircraft mounted FLIR camera was used, combined with target generation and a cloud occlusion model. This enabled the formation of simulated infrared images that could subsequently be processed so as to test algorithms aimed at improving the quality and usability of infrared images.

Chapter 3 Image Pre-Processing and Registration

3.1 Introduction

The previous chapter has given a description of how current infrared camera/image systems work and introduced a simulated camera model for generating synthetic infrared imagery. This chapter will present some pre-processing methods for infrared images and review a fundamental image processing technique – image registration which is essential for aligning images in a sequence. Some pre-processing methods are necessary for infrared images as they have very low contrast and are noisy.

Pre-processing methods including contrast enhancement, dead or saturated pixel removal and detection of previously pre-processed imagery will be presented in Sections 3.2, 3.3 and 3.4 respectively. The source of the difference between images, geometric transformation models and interpolation will be illustrated in Section 3.5. A review of the state of the art in image registration techniques will be discussed in section 3.6.

3.2 Contrast Enhancement

One of the previously mentioned disadvantages of FLIR images is that they tend to be low contrast. Here contrast enhancement method based on thresholding has been proposed. This method enables to view the real infrared images more obvious by increasing the displayed contrast between image brightness levels. This contrast enhancement method was performed on the infrared images by setting thresholding using the standard deviation. Since the real infrared images are so low contrast and it is very difficult to see the features, this simple contrast enhancement method is aimed to improve the visibility of the image in order to give a general view of the image. Therefore, it is used after all the other processing methods and it can give a better view of the image for human eyes.

This contrast enhancement is applied by using standard deviation to eliminate the extremely high or extremely low pixel values in the real infrared image. Standard deviation is a common measure of statistical dispersion, which measures how values are spread out in a data set. The standard deviation of a random variable X is defined as:

$$\sigma = \sqrt{E\left((X - E(X))^2\right)} = \sqrt{E(X^2) - (E(X))^2} \tag{3-1}$$

where $E(X)$ is the expected value of X.

The square root of the sample variance of a set of N values is the sample standard deviation (Kenney and Keeping, 1962):

$$S_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{3-2}$$

where $\bar{x}$ is the mean value of X.

For normally distributed variables, about 68% of the values are within 1 standard deviation of the mean, about 95% of the values are within two standard deviations and about 99.7% lie within 3 standard deviations (Taylor, 1997). Diagram of standard deviation is shown diagrammatically in Figure 3.1. After obtaining a standard deviation of the intensities of all the pixels on the image, the threshold for enhancing the image can be decided. The threshold can be set as 2 to 3 times the standard deviation. In this project, 2.5 times the standard deviation has been chosen as the threshold for enhancing the real image.

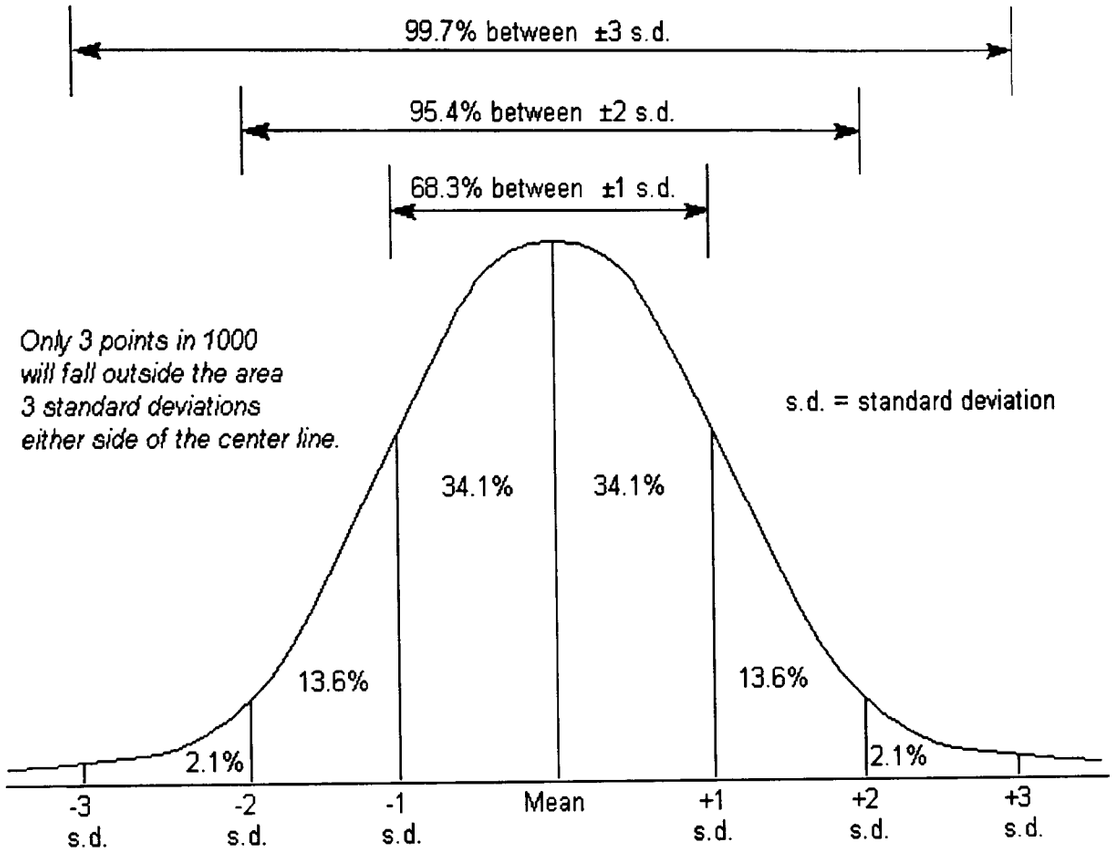
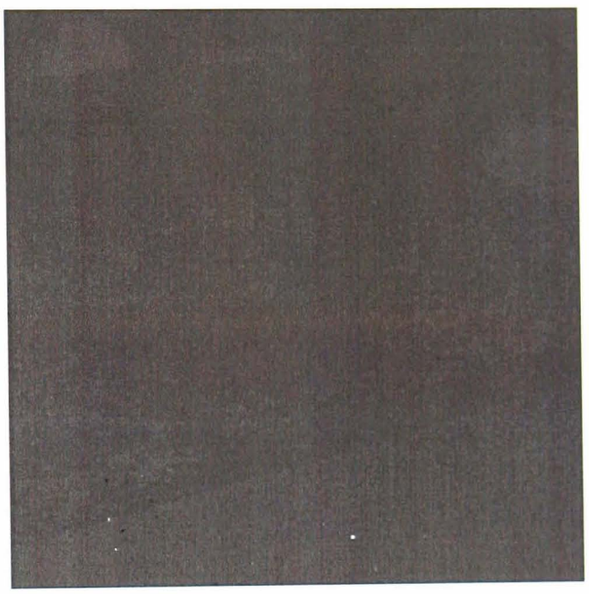


Figure 3.1 Percentages under the normal curve ((Taylor, 1997).

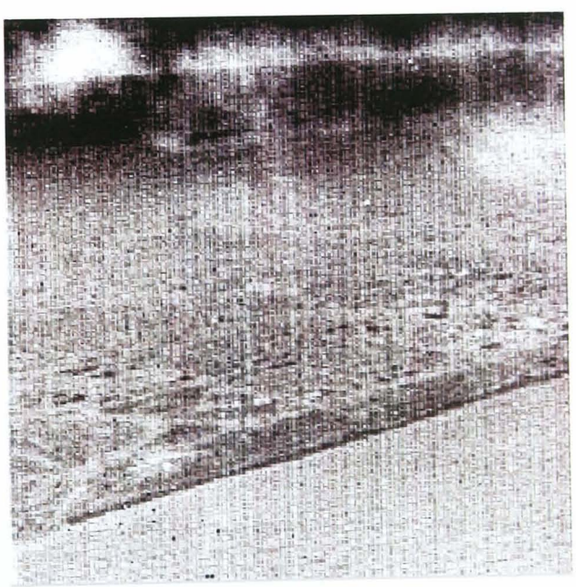
In this investigation, a sequence of 256×256 pixel real infrared images has been used for contrast enhancement testing. The algorithm for enhancing the image is as follows:

1. *For each pixel intensity value* x_i, *calculate the difference* $x_i - \bar{x}$ *between* x_i *and the mean* $\bar{x}$
2. *Calculate the standard deviation* σ *of the whole image*
3. *Set the low threshold as* $\bar{x} - 2.5\sigma$ *and set the high threshold as* $\bar{x} + 2.5\sigma$
4. *If the pixel value is above the high threshold, set the value for that pixel to* $\bar{x} + 2.5\sigma$
5. *If the pixel is below the low threshold, set the value for that pixel to* $\bar{x} - 2.5\sigma$
6. *If the pixel is within the high and low threshold boundaries, then its value remains unchanged*

An example of an original infrared image prior to enhancement is shown in Figure 3.2 (a). The features in the image are very difficult to resolve due to the poor contrast. After processing this image using the standard deviation algorithm, the contrast enhanced image is shown in Figure 3.2(b). The features in the enhanced image are easier to see and detail not visible in the original image is now apparent.

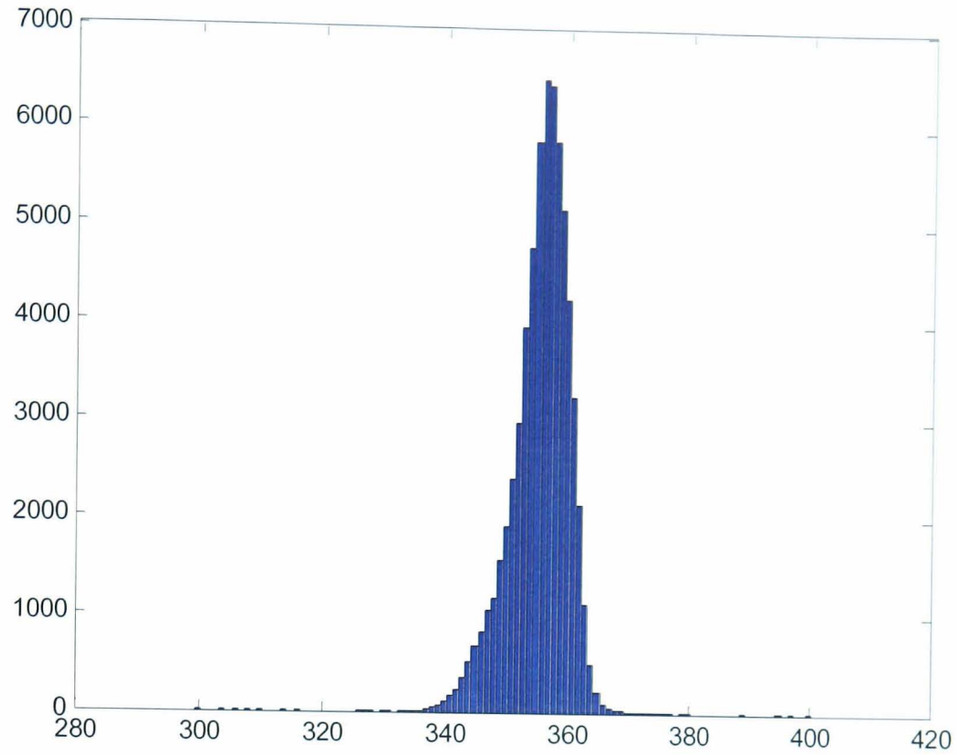


(a)

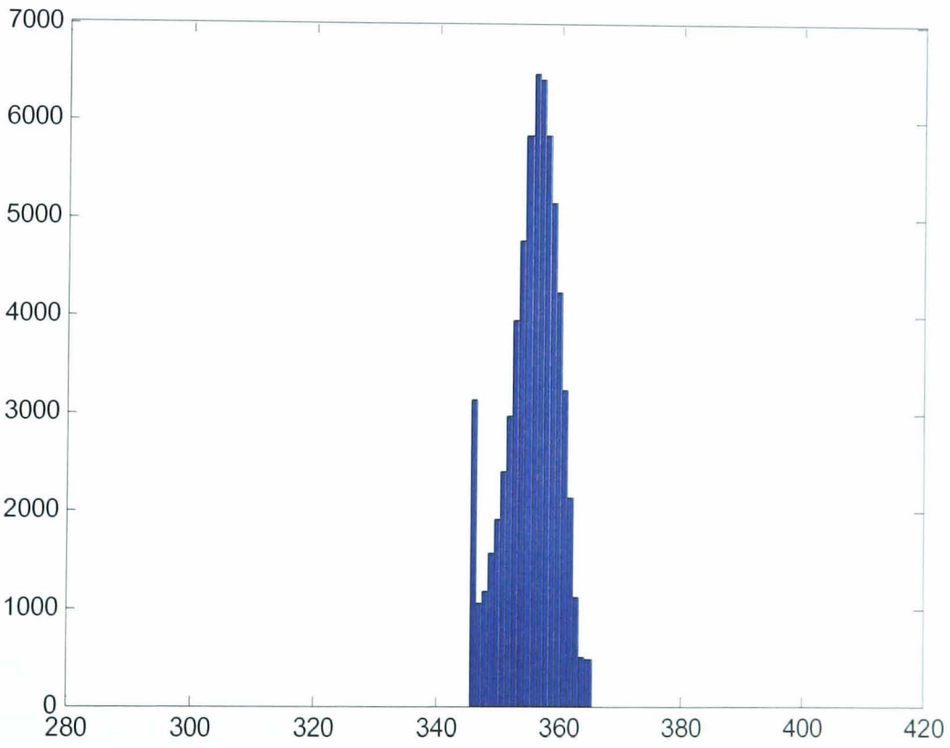


(b)

Figure 3.2 Example of contrast enhancement on a 256x256 infrared image: (a) infrared image prior to contrast enhancement; (b) infrared image after contrast enhancement



(a)



(b)

Figure 3.3 Histogram of the real FLIR images before enhancement and after enhancement: (a) histogram prior to contrast enhancement; (b) histogram after contrast enhancement

The histograms of the real FLIR image before and after enhancement have been shown as Figure 3.3. Along the vertical axis it the intensity value of the pixel. and along the horizontal axis is the number of the pixels in the image have the same intensity value. As shown in Figure 3.3(a), the intensity values of the image pixel have been relatively intense between intensity value 340 and 370. However, there are a few scattered out side this band and it is the reason the original real image have very low contrast to see. Therefore, after enhancement by thresholding the scattered intensity values, the histogram in Figure 3.3(b) shows that the intensity value have been more intense compared with the histogram shown in Figure 3.3(a).

3.3 Dead/Saturated Pixel Removal

Dead or saturated pixels are a common type of noise in infrared images caused by defects in the detector pixel array. The dead/saturated pixels in infrared images can be removed by applying a median filter.

Median-filter-based methods are one of the most popular approaches for dead/saturated noise reduction in image processing. This uses a non-linear smoothing method which replaces the current point in the image by the median of the brightness of the neighbouring pixels (Sonka et al., 2008). An example illustrating a 3×3 median filter is shown in Fig. 3.4.

The filter algorithm screens questionable pixels, that is, those with very high or very low values. The median value is then found by first sorting the pixel values into order and then replacing the suspected dead/saturated pixel with the median value. A disadvantage is that only a small amount of the total information available is used, as effectively this technique simply duplicates a neighbouring pixel value. (Miller, 1994).

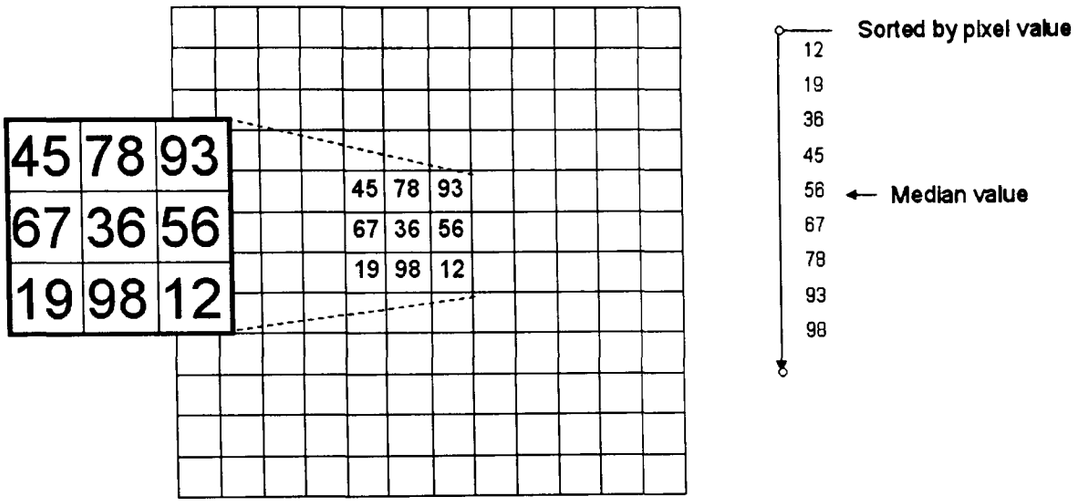


Figure 3.4 An example of illustrating a 3×3 median filter

Some infrared camera systems have a built-in median filter function to remove dead/saturated pixels automatically. Normally in these camera systems, the median filter is applied to the dead/saturated pixels only, rather than the whole image so as to avoid blurring thin lines, sharp corners and edges.

3.3.1 Synthetic Dead/Saturated Image Processing

So as to assess the effect of the median filter in dealing with dead or saturated pixels in infrared images, synthetic images were created with this noise simulated. The median filter algorithm was then applied and the images compared.

A synthetic infrared image with dead/saturated pixel noise is shown in Figure 3.5(a). The result of applying the median filter algorithm to the dead/saturated pixels is shown in Figure 3.5(b). It can be seen that the median filter was effective in removing the dead/saturated pixels from the image and consequently accounting for the simulated sensor defects. This improves visual inspection by observers but does not add any extra information to the image.

(a) (b)

Figure 3.5 Examples of synthetic infrared images: (a) Infrared image with dead/saturated pixels, (b) infrared image after processing by a median filter.

3.4 Detection of Pre-processing

The built in median filter function of some cameras ensures images are more acceptable to human observers, but replaces dead or saturated pixels with information that is not a true representation of the defective pixel. Due to this effect, the application of a median filter to an image can cause problems in subsequent processing, such as when utilising the super resolution technique.

The super resolution technique needs to make use of all of the image pixels and subsequently an error occurs when the value of a processed pixel is not its true value. Real infrared imagery is normally generated with a log, providing further information. However, it is quite common that the real image data provided by industry is lack of the log information. If the dead/saturated pixels are included in this information, it is relatively easy to identify processed pixels and exclude them from further processing. However, if there is a lack of this information, it can be very difficult to identify which pixel has been pre-processed by the median filter and

consequently subsequent post-processing by such things as the super-resolution algorithm is not possible. To determine the pre-processed pixels, a novel detection algorithm has been proposed and tested on both synthetic and real imagery.

The basic idea of the detection algorithm is based on reversing the process of the median filter. If a pixel is dead or saturated, there must be at least two identical pixels in the filter patch equal to the median value after applying the median filter. The proposed algorithm applies a filter the same size as the median filter to the whole image to find out whether the centre pixel value is the same as any other pixel in the patch and is equal to the median value. This is then applied to each frame in a sequence of moving frames. If the intensity value of the centre pixel is the same as the median value and more than one other pixel value in each frame of the sequence, then it is a pre-processed dead/saturated pixel. However, the size of median filter used may be unknown. Therefore, several filters with different sizes can be tested to determine the right size of the filter.

The algorithm for the basic detection algorithm is shown as follows:

1. *For each frame, apply a $n \times n$ filter patch on the whole image ($n = 7,9,11...$)*
2. *If the pixel intensity value is the same as any other pixel and the median value in the filter patch, save its corresponding location to a new dummy image*
3. *Go to new frame. If in the new frame, the pixel does not satisfy condition 2 while it is satisfied in the previous frame, update the dummy image to change the value to zero*
4. *Update dummy image and update the detected pixel numbers through the whole sequence of images. Then plot the frame numbers of the detected pixels*

3.4.1 Pixel Detection on Synthetic Imagery

The detection algorithm was first tested on synthetic imagery which had not been pre-processed with the median filter. An example of a frame of the non processed infrared image used in this test is shown in Figure 3.6(a). The detection algorithm was then applied in 7×7, 9×9 and 11×11 patch filters, shown in Figure 3.6(b), (c) and (d) respectively.

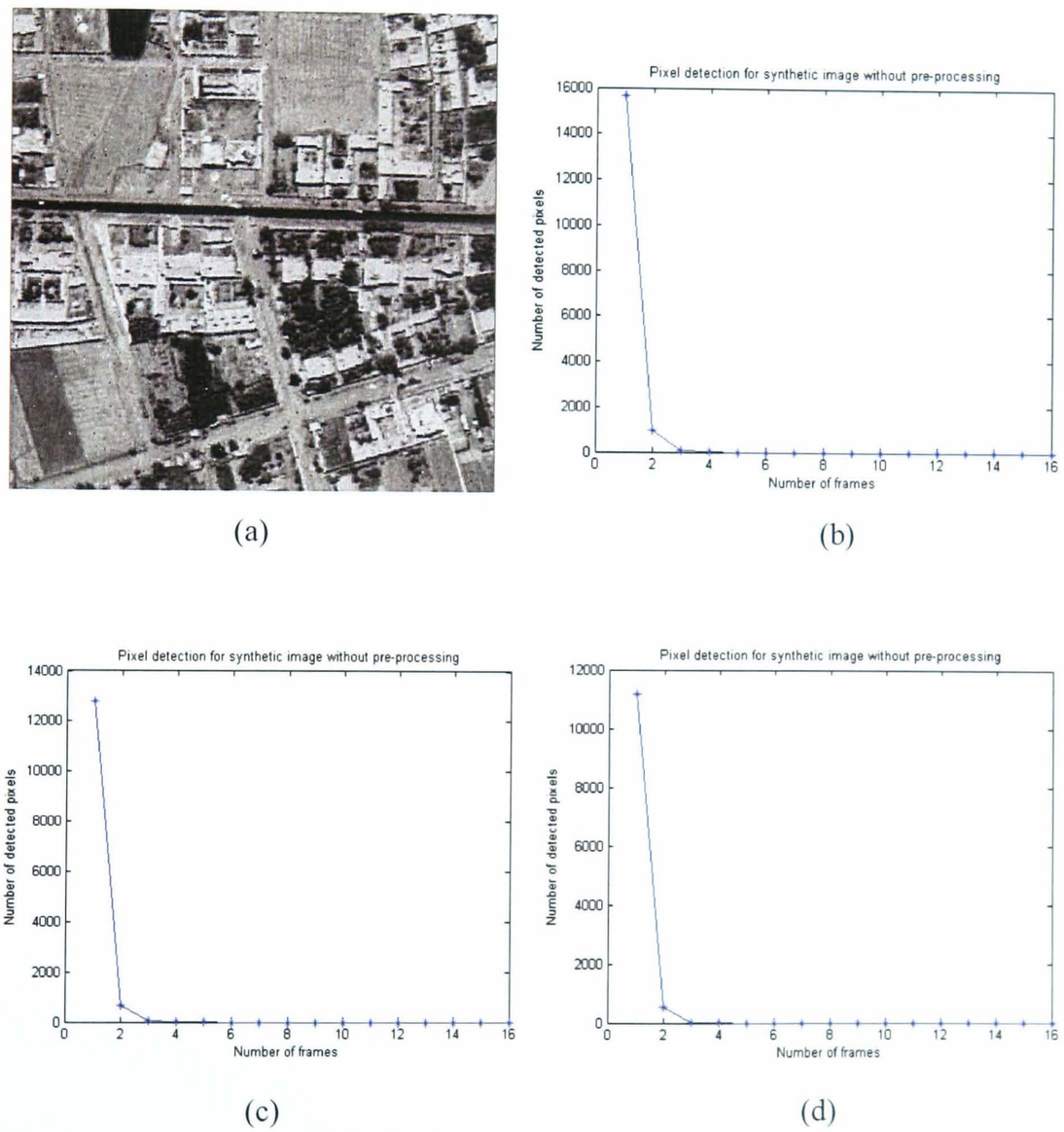


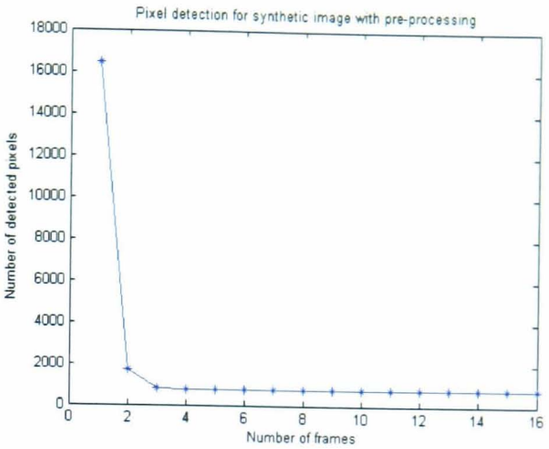
Figure 3.6 (a) An example of synthetic infrared frame without pre-processing, (b) pixel detection result with *7×7* median filter, (c) pixel detection result with *9×9* median filter, (d) pixel detection result with *11×11* median filter.

The results of testing the non pre-processed imagery with the detection algorithm show that no pixels were detected as being pre-processed, which is the expected result. The next stage took a synthetic image that had been pre-processed with a median filter to remove dead/saturated pixels. Figure 3.7(a) shows one frame of the pre-processed infrared image. The detection algorithm was then applied to the image with three different patch filter sizes, 7×7, 9×9 and 11×11. The results of the detection algorithm are shown in Figures 3.7(b), (d) and (e).

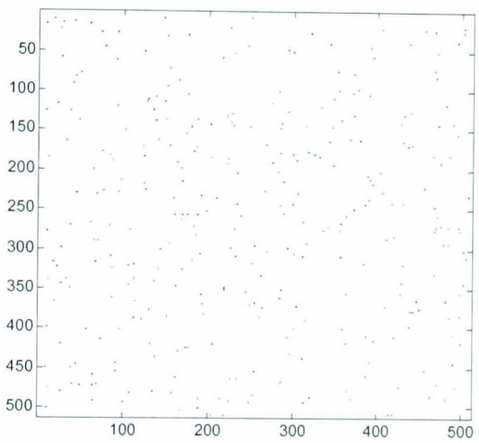
Figure 3.7(b), the results for the 7×7 patch size, shows that the algorithm detected a number of pre-processed pixels, this not only shows that this image has been pre-processed but that the 7×7 patch filter size is the correct one. There are in total 741 pre-processed pixels were detected in Figure 3.7(b). The pixels detected as pre-processed are shown in the dummy image in Figure 3.7(c).

The results for the 9×9 and 11×11 patch sizes do not detect any pre-processed pixels because the patch size for those was not correct. This can be seen in Figures 3.7(d) and (e).

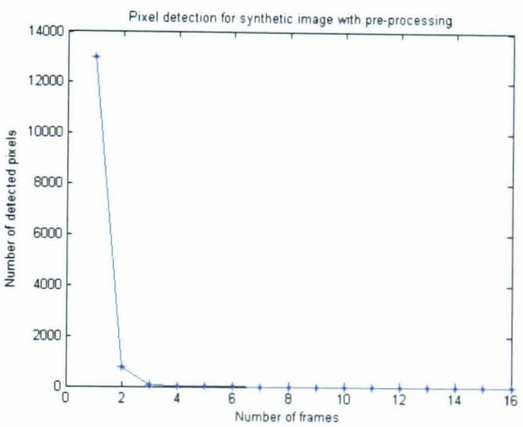
(a)



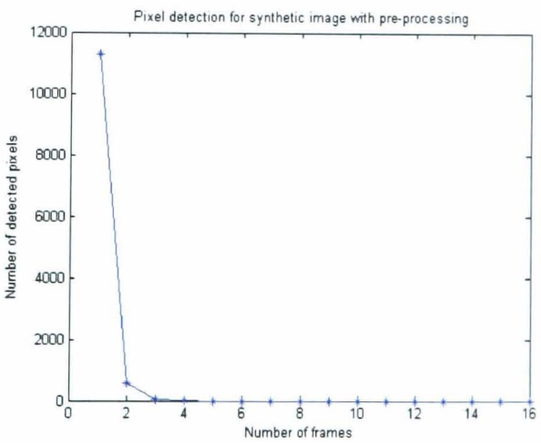
(b)



(c)



(d)



(e)

Figure 3.7 (a) An example of a synthetic infrared frame with pre-processing, (b) pixel detection result with *7×7* median filter, (c) detected dead/saturated pixel on dummy image (d) pixel detection result with *9×9* median filter, (e) pixel detection result with *11×11* median filter.

Overall, the results of detecting the synthetic infrared imagery with and without pre-processing were distinctive. For the synthetic infrared imagery without pre-processing, the test results show zero detected pixels after certain frames, whereas with correct patch size, the test results with pre-processing by the median filter show a positive number of detected pixels. Therefore, for synthetic images with pre-processing, the processed pixels can be identified and their locations can be indicated on the dummy image.

3.4.2 Pixel Detection on Real Infrared Imagery

After testing the detection algorithm on synthetic infrared images, it was applied to real infrared imagery without and with pre-processing. The infrared imagery without pre-processing has been tested first. Figure 3.8(a) shows one frame of real infrared imagery, which has not been pre-processed by the median filter and consequently still contains dead/saturated pixels. The median filter detection algorithm was applied to the image sequence and the results shown in Figure 3.8(b), (c) and (d) for different patch filter sizes.

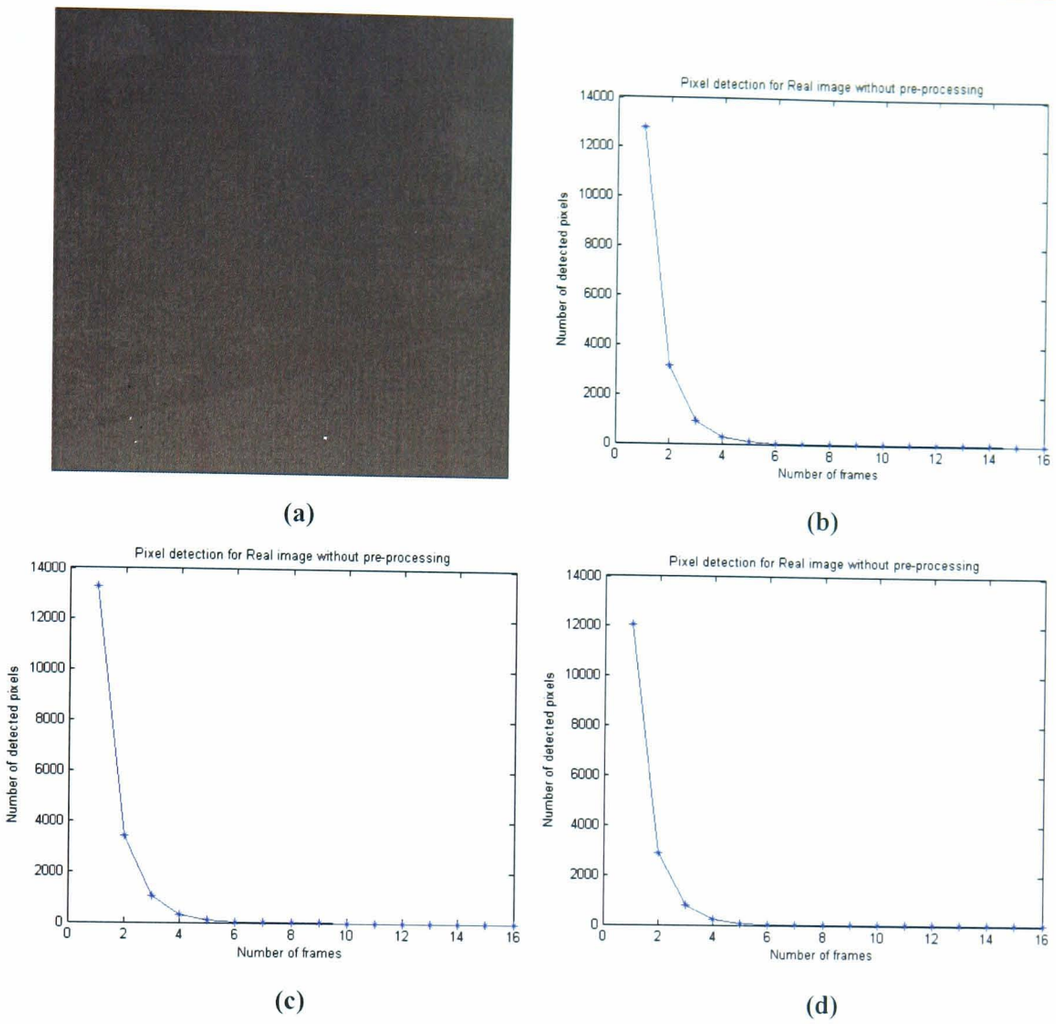
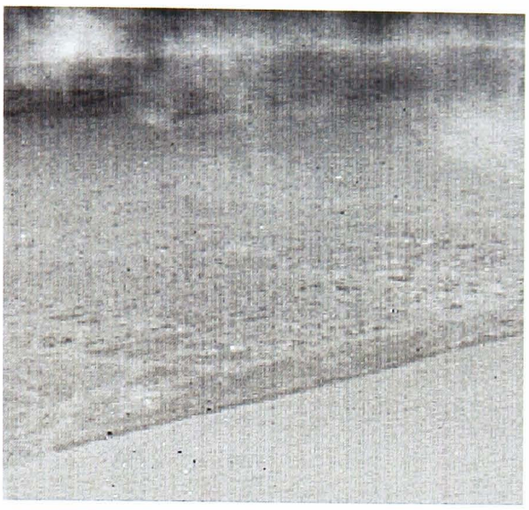


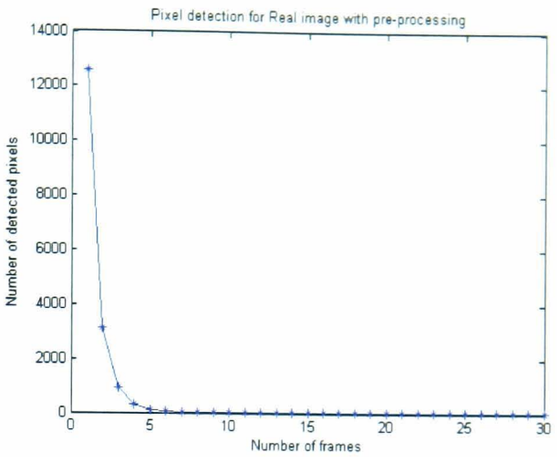
Figure 3.8 (a) An example of real infrared frame without pre-processing, (b) pixel detection result with *7×7* median filter, (c) pixel detection result with *9×9* median filter, (d) pixel detection result with *11×11* median filter.

After applying the median filter detection algorithm with different patch sizes, all the final numbers of detected pixels are zero. This confirms that this infrared image has not been pre-processed by the median filter which is just like the tests on synthetic imagery.

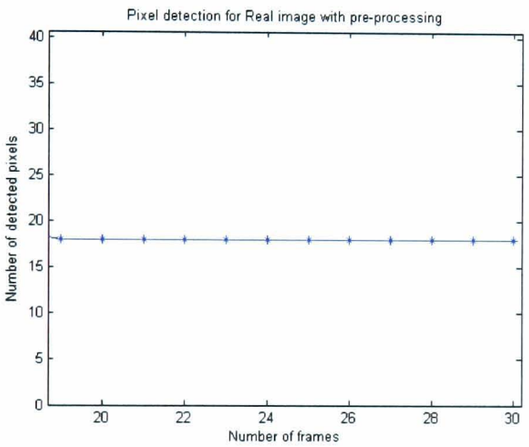
Next the infrared imagery with pre-processing has been tested. Figure 3.9(a) shows one frame of real infrared imagery, which has been pre-processed by the median filter. The median filter detection algorithm was applied to the real imagery and the results shown in Figure 3.9(b), (c) and (d) for different patch filter sizes, 7×7, 9×9 and 11×11.



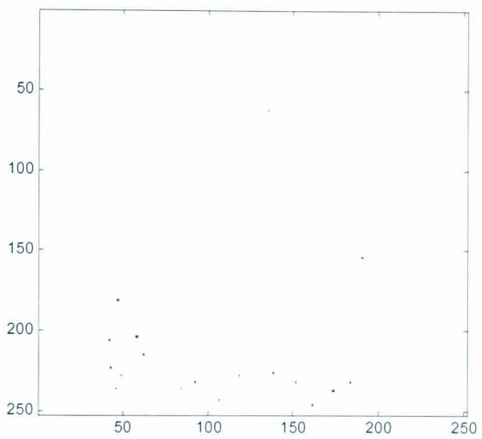
(a)



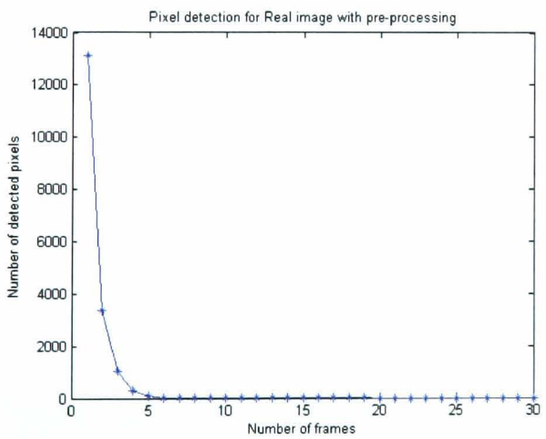
(b)



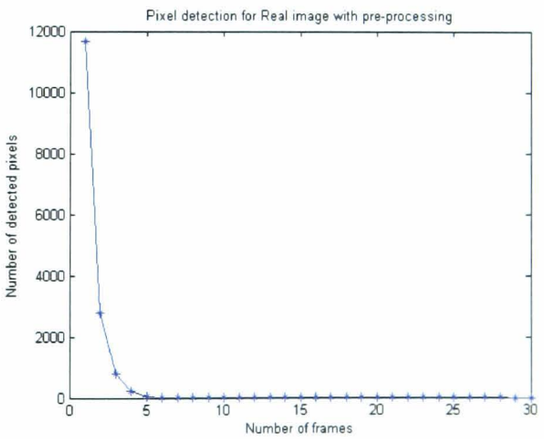
(c)



(d)



(e)



(f)

Figure 3.9 (a) An example of real infrared frame with pre-processing, (b) pixel detection result with *7×7* median filter, (c) enlarged pixel detection result with *7×7* median filter, (d) detected dead/saturated pixel on dummy image, (e) pixel detection result with *9×9* median filter, (f) pixel detection result with *11×11* median filter.

Figure 3.9(b), the results for the 7×7 patch size, shows that the algorithm detected a number of pre-processed pixels in the real images. This not only shows that this image has been pre-processed but that the 7×7 patch filter size is the correct one. As the final number of the detected pre-processed pixels is relatively small in Figure 3.8 (b), the end of the figure has been enlarged as shown in Figure 3.9 (c). The pixels detected as pre-processed are shown in the dummy image in Figure 3.9 (d). The results for the 9×9 and 11×11 patch sizes do not detect any pre-processed pixels because the patch size for those was not correct. This can be seen in Figures 3.9 (e) and (f).

In the detection results, the final number of detected pixels is zero for the 9×9 and 11×11 median filters except for 7×7 median filter. The final number of detected pixels by 7×7 median filter is 18 which indicates that the real infrared imagery has been pre-processed.

3.5 Overview of Image Registration

In the last few sections, some pre-processing tasks have been performed in order to render the resulting infrared images more suitable for the further processing. Next, a fundamental technique for aligning moving image sequences will be discussed. This is a requirement for FLIR images so as to account for the moving camera and changes in background.

Image registration is a basic imaging process of matching two or more images taken at different times, or from different viewpoints, or from different sensors. It is a process of aligning the pixels in one image to the corresponding pixels of a second image, finding the optimal transformation between images.

Image registration has applications in a number of different image processing

research areas, including computer vision and pattern recognition, medical image analysis and remotely sensed data processing (Zitova and Flusser, 2003). Precise image registration is a fundamental requirement for many image processing tasks such as generating a high resolution image from a sequence of images, fusing two or more images into a composite one that reveals information not easily accessible within individual images, or detecting a change between images by comparing image differences.

For image registration processing, it is necessary to find an appropriate mathematical transformation model relating pixel coordinates in one image to pixel coordinated in another. Then, the correct alignments of various pairs of images can be estimated. A representation of transformations of rotation ϕ, and t_x and t_y translations between input image and reference images is shown in Figure 3.10. The origin is the centre of the template image. This coordinate system will be used throughout the thesis.

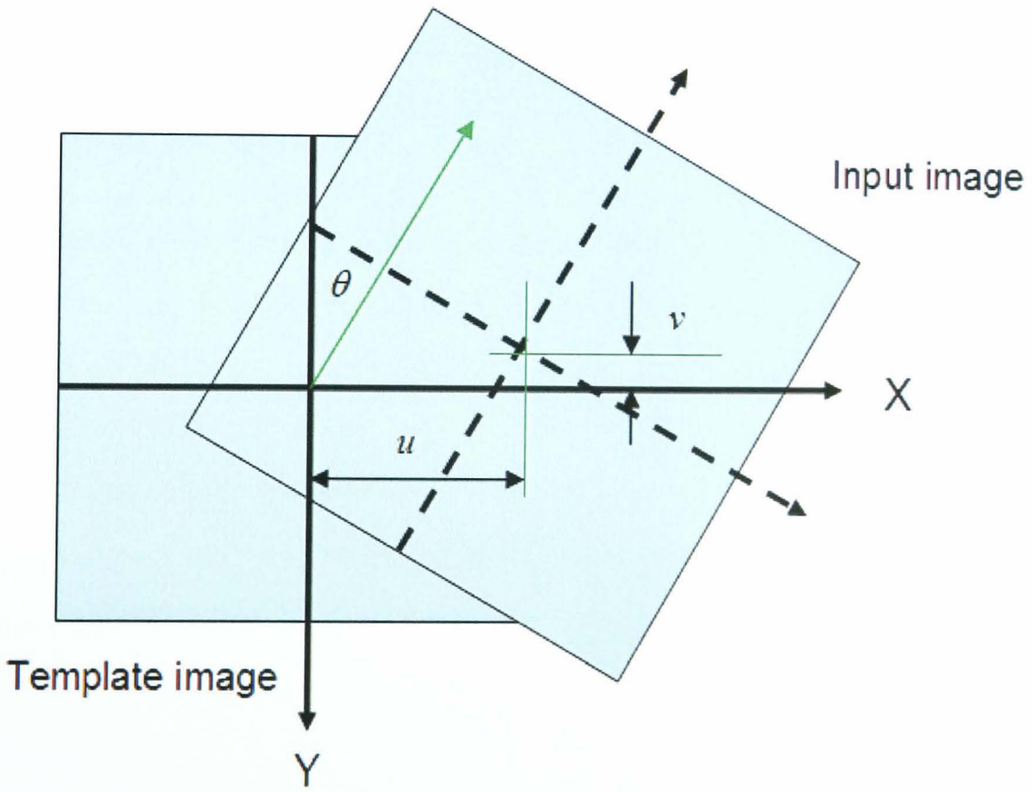


Figure 3.10 An example of transformations between input image and template images rotation θ and translations t_x and t_y

3.5.1 Differences between Images

There are two main differences to consider when registering images in sequences, the first is caused by changes in camera position and the second is caused by such factors as lighting changes and differences between sensors.

Changes caused by camera position and pose cause the images to be spatially misaligned, which includes relative translations, rotation, scale, and other geometric transformations in relation to each other. These are considered as spatial differences. In practice, changing the orientation or parameters of the imaging sensor can lead to spatial transformations.

The second type of difference, which is attributed to factors such as lighting changes, using different types of sensors, using similar sensors but with different sensor parameters, object movement, or scene changes cannot be modelled by a spatial transform alone. This type of difference is therefore considered to be a non-spatial difference.

The independent movement of objects within the scene can cause changes between images. There are generally three scenarios that lead to occlusions: part of a finite image moves out of the image frame, new data enters the image frame of a finite image, or an obstruction comes between the imaging sensor and the object being imaged. These non-spatial differences cannot be removed by registration and in fact make registration more difficult since there is no longer an exact match between the two images, even after the application of a spatial transformation. Consequently, non-spatial differences make it more difficult to achieve the accurate registration of the images.

For FLIR imaging, spatial transformations are common due to the camera being mounted on a moving platform. It is also possible that a scale change can occur due

to variation in distance between sensor and scene. In FLIR imaging, changes in scene lighting are not a large issue compared to visible band imaging, however changes in background temperature do introduce non-spatial issues which affect the brightness of FLIR images. This situation tends not to occur in short timescales. In practice, real FLIR images can be collected at different times of day, leading to temperature differences causing variations in image brightness.

Some issues related specifically to registration of airborne FLIR imaging applications are clouds and moving objects. Clouds cause occlusions between the detector and the background, whereas independently moving objects cause differences in the background between frames in an image sequence.

After aligning the differences caused by the movement of the camera, removing the occlusions and reducing noise, the differences between the changes of the images can be preserved for detection purposes.

3.5.2 Image Registration Processes

When applying image registration techniques on a pair of images, one of the images, known as the template image, is kept unchanged and is used as a reference. The other image, known as the sensed image, is registered with respect to this template image. The parameters for the transformation are determined by finding how to map the sensed image to the template image. The key stage in image registration is to find the geometric transformations between images in a sequence.

Given two images I_1 and I_2, where $I_1(x,y)$ and $I_2(x,y)$ represent their respective intensity values, the transformation between I_1 and I_2 can be expressed as:

$$I_2(x_1, y_1) = g(I_1(T(x_1, y_1))) \tag{3-3}$$

where T is a 2D spatial coordinate transformation matrix and g represents a 1D intensity or radiometric transformation (Brown 1992). The pixel (x_1, y_1) in I_1 is mapped to a new position (x_2, y_2) in I_2 as:

$$(x_2, y_2) = T(x_1, y_1) \tag{3-4}$$

The transformation matrix between the template image and sensed image can be calculated by mapping the co-ordinate of the sensed image pixel to the point in the template image. An illustration of the whole region with spatial transformations based on point to point is shown in Figure 3.11.

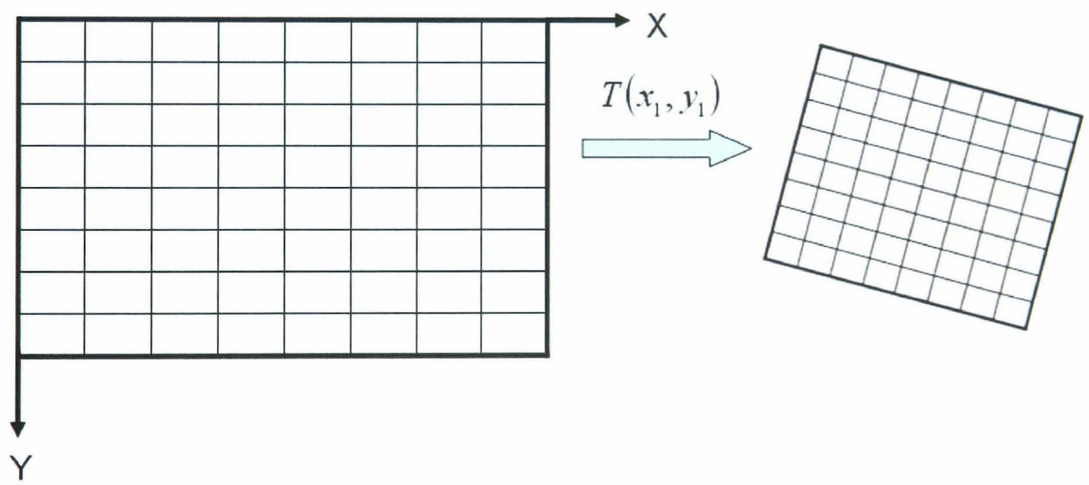


Figure 3.11 An illustration of the whole region transformation

There are different image registration methods that can be used depending on the differences in images and imaging conditions. Generally, the image registration can be divided as feature-based and intensity-based. The process of feature-based image registration can be considered as a selection of these four steps: feature selection; feature matching; transformation model estimation; and image transformation and

brightness interpolation (Zitova and Flusser, 2003). For intensity-based image registration method, the image intensity is used to find the correspondence.

In feature selection, a feature is a salient or distinctive object which can be manually or automatically detected. These features, also known as control points, can be represented by distinctive points, line endings and the centres of gravity of regions. If features are detected, then the correspondence between the features can be established, and consequently the transformation between images can be found by a similarity metric.

The transformation can be made in a forward or backward way. The forward registration requires all of the transformation to be registered every time a new image is obtained. In backward method, the registered image data is represented using the same coordinate system as the template image. Therefore, the backward approach is normally used as it is simpler to implement (Zitova and Flusser, 2003).

The parameters of transformation model estimation are computed either by the established correspondence between features or directly by image intensity. Then, the sensed image is transformed by the transformation estimation. After transformation, the coordinate of output point does not need be integer as the output position does not necessarily match the digital grid (pixellated image points). To find the point in the digital raster which matches the transformed point and determine its brightness value, the brightness is usually computed as an interpolation of the brightness of several points in the neighbourhood. The final output image is the registered image.

3.5.3 Transformation Models

Image transformation models are the most commonly used mathematical ways to describe image registration. The simplest and most used general geometric

transformations are rigid, affine and projective (Brown, 1992). These transformations can be categorised based on the geometric transforms for planar surface elements as translation, rotation, scaling, stretching, and shearing as shown in Figure 3.12.

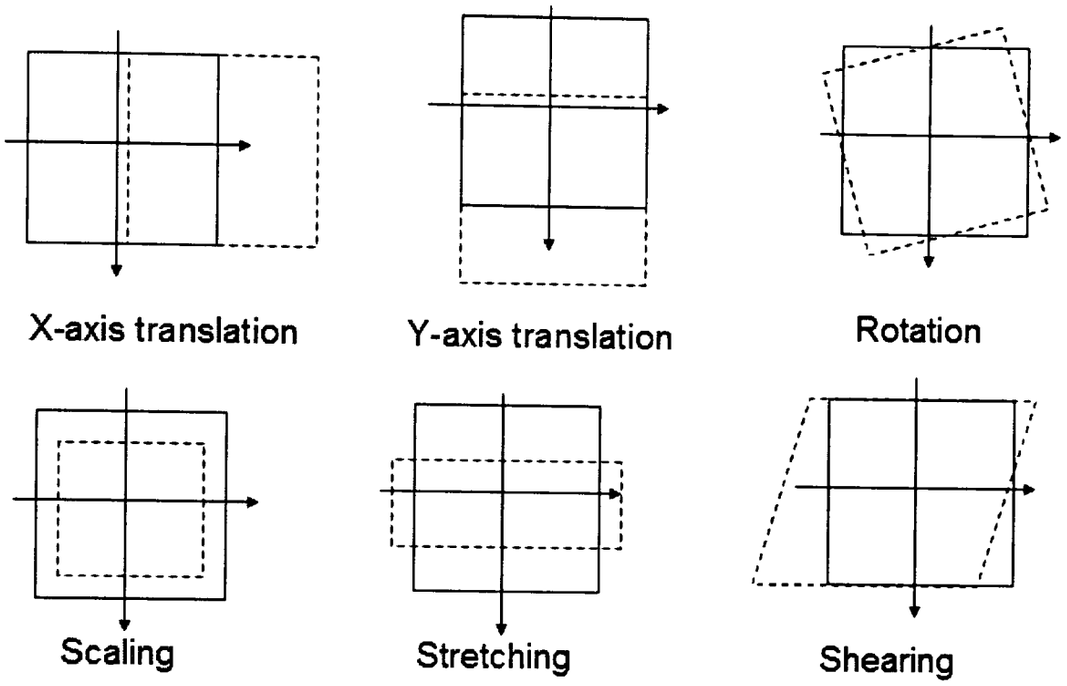


Figure 3.12 Basic 2D planar transformations

Translations and rotation are caused by different orientation of the imaging sensor while scaling is the effect of a change in aspect ratio. Sensor distortion or different viewing angles may cause stretching or shearing (Chalermwat, 1999).

3.5.3.1 Rigid Transformation

Rigid transformation generally occurs by the movement of the camera. Rigid transformation is composed of a combination of a rotation, a translation and a scale change. The rigid transformation can be expressed mathematically as follows:

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + s \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \tag{3-5}$$

where (x_1, y_1) point in the template image, (x_2, y_2) are corresponding points after transformation, t_x, t_y are the shifts along X and Y axis, s and θ are scale factor and rotation angle respectively.

3.5.3.2 Affine Transformation

Affine transformations are more general than rigid transformation. Affine transformations are linear in the sense that they map parallel lines onto parallel lines. An affine transformation is composed of a rigid transformation, a shear, and an aspect-ratio change (Le Moigne et al., 2002). Shear is generally caused by the distortion of the lenses of the imaging system. Shear is a distortion that acts either along the *x*-axis, $Shear_x$ or along the *y*-axis, $Shear_y$ and can be represented as:

$$Shear_x = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \; Shear_y = \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \tag{3-6}$$

A shear in the *x/y* axis transforms the *x/y* coordinate into a linear combination of both *x* and *y* coordinates. Aspect ratio is defined as the numerical ratio of image width to height (Le Moigne et al., 2002). The general 2D affine transformation can be expressed as (Brown, 1992):

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \tag{3-7}$$

where (x_1, y_1) point in reference image, (x_2, y_2) are corresponding points after transformation, the *a* terms are constants.

3.5.3.3 Projective Transformation

Projective transformation is more general and accounts for the distortion which occurs when a 3D scene is projected onto a 2D plane. It maps straight lines onto straight lines and can be represented as follows (Brown, 1992):

$$x_i = \frac{a_{11}x_p + a_{12}y_p + a_{13}}{a_{31}x_p + a_{32}y_p + a_{33}} \tag{3-8}$$

$$y_i = \frac{a_{21}x_p + a_{22}y_p + a_{23}}{a_{31}x_p + a_{32}y_p + a_{33}} \tag{3-9}$$

where (x_p, y_p) is a point in the template image, (x_i, y_i) are corresponding points after transformation, the a terms are constants.

For airborne FLIR imagery, images are captured from a moving platform and the aircraft is so high above the ground that any terrain effects are negligible. Based on this fact, the ground can be considered as approximately flat. Consequently, a two-dimensional affine or projective transformation is enough to be used to describe the transformation between frames.

3.5.4 Brightness Interpolation

After transformation, the registered image is allocated on non-integer co-ordinates. Therefore, brightness interpolation is needed to determine the integer values on the integer grid for each pixel value in the registered image. Interpolation is performed by an interpolation kernel on the registered image. Generally, the simpler the interpolation the greater is the loss in geometric and photometric accuracy (Milan et al., 2008). Parker et al. (1983) and Lehmann et al. (1999) have given a detailed investigation and comparison of interpolation methods for 2D images. There are

three main interpolation methods: nearest neighbour, bilinear and bi-cubic (Abramowitz and Stegun, 1972).

3.5.4.1 Nearest Neighbour Interpolation

Nearest-neighbourhood interpolation assigns to the point (x, y) the brightness value of the nearest point in the discrete raster, which is illustrated in Figure 3.13. The output pixel is assigned the value of the pixel that the point falls within. No other pixels are considered. The position error of the nearest-neighbourhood interpolation is at most half a pixel; the error can be perceptible on objects with straight-line boundaries that may appear step-like after the transformation, as shown in Figure 3.14.

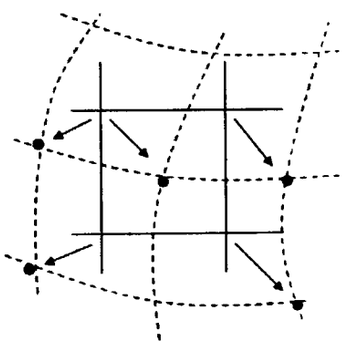


Figure 3.13 Nearest neighbourhood interpolation. (The discrete raster of the original image is depicted by the solid line, dashed lines show the how the inverse planar transformation maps the raster of the output image onto the input image) (Milan et al., 2008)

Figure 3.14 Nearest neighbour interpolation causing stepping at straight line boundaries (Zitova and Flusser, 2003)

3.5.4.2 Bilinear Interpolation

In bilinear interpolation the output pixel value of point (x,y) is a weighted average of pixels in the nearest 2-by-2 neighbourhood and assumes that the brightness function is linear. Fig. 3.15 shows the illustration of bilinear interpolation.

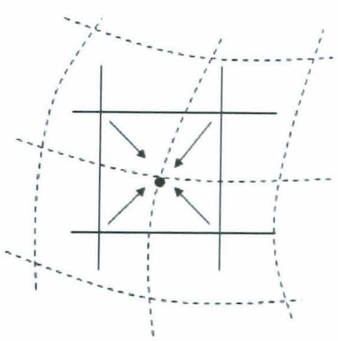


Figure 3.15 Bilinear interpolation. (The discrete raster of the original image is depicted by the solid line and the dashed lines show the how the inverse planar transformation maps the raster of the output image into the input image) (Milan et al., 2008)

Given location (x, y) and assuming u is the integer part of x and v is the integer part of y, the intensity at (x, y) is estimated from the intensities at (u, v), $(u+1, v)$, $(u, v+1)$, $(u+1, v+1)$. This intensity at (x, y) is computed from the linear interpolation as follows (Abramowitz and Stegun, 1972, Milan et al., 2008):

$$I(X,Y) = W_{u,v} I(u,v) + W_{u+1,v} I(u+1,v) + W_{u,v+1} I(u,v+1) + W_{u+1,v+1} I(u+1.v+1) \quad (3\text{-}10)$$

where

$$W_{u,v} = (u+1-x)(v+1-y),$$

$$W_{u+1,v} = (x-u)(v+1-y),$$

$$W_{u,v+1} = (u+1-x)(y-v),$$

$$W_{u+1,v+1} = (x-u)(y-v).$$

Bilinear interpolation can cause blurring and a small decrease in resolution due to its averaging nature. The problem of step-like straight boundaries evident in nearest neighbour interpolation is reduced. An example of bilinear interpolation is demonstrated in Fig. 3.16.

Figure 3.16 Bilinear interpolation with reduced stepping at straight line boundaries compared to nearest neighbour interpolation (Zitova and Flusser, 2003)

3.5.4.3 Bi-Cubic Interpolation

Bi-cubic interpolation improves the model of the brightness function by approximating it locally by a bi-cubic polynomial surface. Sixteen neighbouring points are used for interpolation, and the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighbourhood. Bi-cubic interpolation does not suffer from the step-like boundary problem of nearest-neighbourhood interpolation, and copes with linear interpolation blurring relatively well and consequently preserves fine details in the image well. It is often used in raster displays that enable zooming with respect to an arbitrary point, and example of an image processed with bi-cubic interpolation is shown in Figure 3.17.

Figure 3.17 Bi-cubic interpolation (Zitova and Flusser, 2003).

3.5.4.4 Other Methods of Interpolation

Apart from the most common interpolation methods mentioned above, there are various other interpolation methods such as *sinc* functions, quadratic, Lagrange and Gaussion kernels (Lehmann et al., 1999). The *sinc* function is the optimal interpolation technique but is hard to implement due to its infinite extent of *sinc* function. Cubic interpolation is more suitable when the geometric transformation involves a significant enlargement of the sensed image.

The high order methods such as Lagrange and Gaussian kernels have more complex computations. Overall there is a trade-off between accuracy and computational complexity, in practice the bilinear interpolation method appears to be the more commonly used approach in general cases(Zitova and Flusser, 2003).

3.6 Reviews on Image Registration Techniques

Image registration has developed rapidly during the last few decades. A large amount of research has been done on automatic registration techniques. Most literature on image registration techniques has studied on visible images and medical images (Brown 1992, Maintz and Viergever, 1998, Zitova and Flusser, 2003). There are a few papers about registering visible image with infrared image which is considered as multimodal registration (Bulanon et al., 2009, Bilodeau et al., 2011). However, very few articles are about image registration on FLIR imagery specifically. Due to the different characteristics of FLIR imagery, those algorithms that performed well on visible band image and medical images may not be suitable for FLIR imagery. A general review of various image registration methods will provide a basic idea of suitable registration algorithm for FLIR images.

Image registration methods can be categorised with respect to various criteria such as spatial domain or frequency domain, area or features based, local or global, interactive, semi-automatic or automatic, and monomodal or multimodal. Here, image registration methods are categorised as area-based and feature-based methods (Fonseca and Manjunath, 1996, Maintz and Viergever, 1998, Zitova and Flusser, 2003, Haralick and Shapiro, 1993 and Pratt, 1991).

3.6.1 Area Based Methods

The group of area-based matching techniques is the most widely used method due to its relatively simplicity. Area-based registration methods achieve registration by maximising a similarity measure based on the intensity values of the two images. Consequently, they are generally sensitive to the intensity changes, noise disturbance, varying illumination or different sensor types (Zitova and Flusser, 2003).

The advantage of area-based methods is that very accurate registration can be achieved since all of the available data is used; however this is at the cost of very high computational complexity. Classic area-based methods can be generally divided into three groups: correlation based methods; Fourier transform based methods; mutual information based methods.

3.6.1.1 Correlation-based Methods

Anuta (1969) proposed a technique for registering a pair of images. By forming a cross correlation measure between the images, the location of the sensed image can be determined by the maximum correlation value. This method is also known as the correlation coefficient.

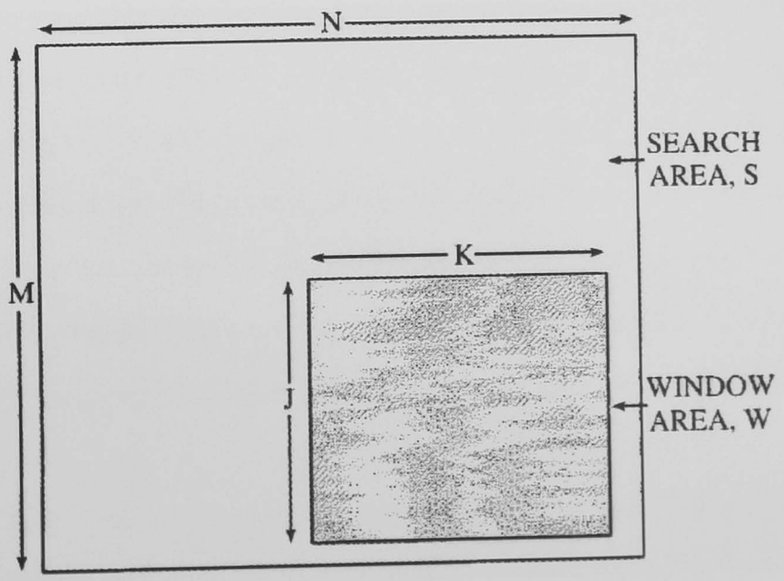


Figure 3.18 Cross correlation search and window areas (Anuta, 1969)

When applying this technique in two dimensions, with $I_1(x,y)$ and $I_2(x,y)$ representing two discrete images to be registered, the cross correlation measure is defined as (Anuta, 1969):

$$C(t_x,t_y)=\frac{\left[\sum_{j=1}^{J}\sum_{k=1}^{K} I_1(x,y)\cdot I_2(x-t_x,y-t_y)\right]}{\left\{\left[\sum_{j=1}^{J}\sum_{k=1}^{K} I_1{}^2(x,y)\right]^{\frac{1}{2}}\cdot\left[\sum_{j=1}^{J}\sum_{k=1}^{K} I_2{}^2(x-t_x,y-t_y)\right]^{\frac{1}{2}}\right\}} \tag{3-11}$$

where (x,y) are indices in a $J\times K$ point window area W that is located within $M\times N$ point search area S as shown in Figure 3.17.

The correlation function $C(t_x,t_y)$ must be computed for all of the search area to determine the maximum value. When the correlation value is a maximum, the two overlapped images will be matched since the similarity between the template image and the other image will be a maximum. There are two basic problems with this simple correlation measure. One is that the maxima of the cross correlation function may be very broad which makes it difficult to detect the peak and the other is that image noise may mask the peak correlation. Pratt (1974) has extended the basic concept of cross correlation by involving linear spatial pre-processing of images before applying the cross correlation. This method measures the cross correlation between the images after they convolve with a spatial filter. The reason to using the spatial filter is to maximise the correlation peak ratio, so as to sharpen the detected correlation peak which improves the detection ability of image misalignment. The deficiency of this method is that it requires a great amount of computation.

Later on, Lewis (1995) described a normalised cross correlation (NCC) method between an image and a template to overcome a few drawbacks of the basic cross correlation method, such as the range of cross correlation being dependent on both

the size of the template image, and the template and sensed image amplitudes. Correlation is normalised to avoid contributions from local image intensities. The normalised cross correlation (NCC) method can be expressed as (Lewis, 1995):

$$C(u,v)=\frac{\sum_{j=1}^{J}\sum_{k=1}^{K}[I_1(x,y)-\mu_1][I_2(x-u,y-v)-\mu_2]}{\left\{\left[\sum_{j=1}^{J}\sum_{k=1}^{K}I_1^{\,2}(x,y)-\mu_1\right]^2\cdot\left[\sum_{j=1}^{J}\sum_{k=1}^{K}I_2^{\,2}(x-u,y-v)-\mu_2\right]^2\right\}^{\frac{1}{2}}} \tag{3-12}$$

where (x,y) are indices in a $J\times K$ point window area W that is located within $M\times N$ point search area S, μ_1 is the mean value of image I_1 and μ_2 is the mean value of image I_2.

Normalised cross correlation measures similarity by computing global statistics such as mean and variance, and it performs well if the two images are similar in nature with an underlying linear relationship between the image intensities (Cole-Rhodes et al., 2003). There are generally a few drawbacks of NCC. Firstly, NCC is sensitive to noise in the images as it makes use of all the intensity value of the images. Secondly, NCC method is mainly used for images misaligned with small rigid or affine transformations due to the fact that the computational costs increase quickly with the growth of the number of transformations (Brown, 1992). Thirdly, for the measurement to be reliable the displacement has to be greater than the mean error of the image to be registered (Debella-Gilolow and Kääba, 2011). Fourthly, the precision of NCC is limited to one pixel. However, when NCC is used combined with optimisation method, it can achieve sub-pixel accuracy (Cole-Rhodes et al., 2003).

Sarvaiya et al. (2009) describe medical image registration by template matching based on Normalized Cross-Correlation (NCC) using Cauchy-Schwartz inequality.

The algorithm does the template matching and uses the Cauchy-Schwartz's inequality to simplify the procedure. The template matching experiments were only conducted on one greyscale brain image. According to their results, although maximum cross correlation value of image with noise is not as high as the value of image without noise, it still can be used for matching.

The main disadvantage of cross correlation methods is that they lack computational efficiency. A sequential similarity detection algorithm (SSDA) related to cross correlation was proposed by Barnea and Silverman (1972). This method uses a sequential search approach and a computationally simpler distance measure than the cross correlation. It accumulates the sum of absolute differences of the image intensity values as follows:

$$E(t_x,t_y)=\sum_x\sum_y\left|I_1(x,y)-I_2(x-t_x,y-t_y)\right| \tag{3-13}$$

It then applies a threshold criterion. If the accumulated sum exceeds the given threshold, the candidate pair of windows from the template and sensed image is rejected and the next pair is tested. In comparison with cross correlation methods, this method is much simpler computationally since it only computes the absolute differences between the pixels in the two images. Although the sequential methods improve the efficiency of the similarity measure and search, it is likely to be less accurate than the cross correlation method.

3.6.1.2 Fourier based Methods

Fourier methods are frequently preferred over correlation-based methods when there is a need to accelerate the computational speed. Instead of representing the problem in the spatial domain, the Fourier transform represents the image in the frequency domain. It can be computed efficiently for images, especially for large images, using

the two-dimensional Fast Fourier Transform (FFT).

The primary idea of the Fourier method is based on its shift property. The Fourier transform of the correlation of two images is the product of the Fourier transform of one image and the complex conjugate of the Fourier transform of the other (Brown, 1992). This is an alternative way to compute the correlation between images. Given two images I_1 and I_2 which differ only by a displacement (t_x, t_y):

$$I_2(x, y) = I_1(x - t_x, y - t_y) \tag{3-14}$$

their corresponding Fourier transforms, F_1 and F_1 are:

$$F_2(\omega_x, \omega_y) = \exp\{-j(\omega_x t_x + \omega_y t_y)\} F_1(\omega_x, \omega_y) \tag{3-15}$$

The cross-power spectrum of the two images are defined as:

$$\frac{F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)}{|F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)|} = \exp j(\omega_x t_x + \omega_y t_y) \tag{3-16}$$

where F^* is the complex conjugate of F, the phase of cross-power spectrum is equivalent to the phase difference between images. This method is also called phase correlation method (Brown, 1992).

The Fourier method searches for the optimal match in the frequency domain. They are specifically well suited for images with low frequency or frequency-dependent noise, while they are not suitable for images with frequency-independent noise or more general transformations (Zitova and Flusser, 2003). Although the Fourier method can only be used for images with translations, it can be extended to handle

images which have been both shifted and rotated with respect to each other.

If the images are transformed into a polar or a log-polar Fourier grid, rotations and scaling can be reduced to translations in these representations and can easily be estimated by using the phase correlation algorithm. Several papers have discussed various methods for translation with rotation or scaling by the Fourier method based on translation-invariant theory (De Castro and Morandi, 1987, Reddy and Chatterji, 1996, McGuire , 1998, Vandewalle et al., 2006, Kim et al., 2008).

De Castro and Morandi (1987) proposed a method based on FFT for both translations and rotations. They present a two-step process to first determine the angle of rotation and then determine the translational shift. Rotation is invariant to the Fourier transform which means that the rotated angle of an image is the same rotated angle of the Fourier transform of that image. If the angle is unknown, the phase of the cross-power spectrum is computed as a function of the rotation angle ϕ and the equation can be simplified by using polar coordinates (r,θ). Consider two images are related by rotation only. After resampling the images into polar coordinates, the template and sensed image can be expressed in polar coordinates as

$$\tilde{I}_1(r,\theta) = I_1(r\cos\theta, r\sin\theta) \tag{3-17}$$

$$\tilde{I}_2(r,\theta) = I_2(r\cos\theta, r\sin\theta) \tag{3-18}$$

The relationship between two images in polar coordinates can be expressed as

$$\tilde{I}_2(r,\theta-\phi) = \tilde{I}_1(r,\theta) \tag{3-19}$$

In Fourier domain, the phase correlation can be expressed as

$$G(r,\theta,\phi)=\frac{F_1(r,\theta)F_2^*(r,\theta-\phi)}{\left|F_1(r,\theta)F_2^*(r,\theta-\phi)\right|} \tag{3-20}$$

When at the true angle of rotation, the effect of rotation can be removed from the sensed image. Then G should only indicate the value that is expected for the sensed image which has only been translated. Therefore, by searching the angle ϕ which is the closest to maximum value, the rotation angle can be determined. Their results on synthetic and real 2D visible digital image show that the algorithm is robust to correlated noise and time varying illuminations.

Reddy and Chatterji (1996) have discussed an extension of phase correlation technique to cover translation, rotation and scaling. When without considering scale changes, the rotation can be represented in polar coordinates as a translational displacement. If (x,y) in image I_1 is scaled to $\left(\frac{x}{a},\frac{y}{a}\right)$ in image I_2 and with rotation angle ϕ, by transforming images to Fourier domain and in polar representation are related by

$$F_1(r,\theta)=F_2(r/a,\theta-\phi) \tag{3-21}$$

Then by converting into logarithm scale, the scaling can be reduced to a translational movement:

$$F_2(\log r,\theta)=F_1(\log r-\log a,\theta-\phi) \tag{3-22}$$

where log is based on natural logarithm base. Reddy and Chatterji (1996) also have

tested the method on different set images with different amount of noise. The matching for translation and rotation is quite accurate in the presence of noise. However, it is not as accurate for scale change.

These polar-log transform methods operate on the translation-invariant Fourier space first, and then convert to polar-logarithmic coordinates so that rotation and scale effects appear as translational shifts along orthogonal θ and $\log r$ axes. In polar-log space, the normalised correlation coefficient of I_1 and I_2 as a function of shift along these axes is maximised at the coordinate $(-\phi, -\log a)$. The disadvantage of these methods is that by using discrete images instead of continuous ones, some sampling errors occur between the two images and resampling the Fourier magnitudes on the polar-log grid reduces the accuracy of the estimation of rotation and scaling. To enhance accuracy, a few papers introduced new sampling schemes and algorithms which reduce the inaccuracies introduced by resampling the magnitude of the FT on the polar-log grid (Keller et al., 2005, Pan et al., 2009, Tzimiropoulos et al., 2010).

Tzimiropoulos et al. (2010) proposed a robust correlation-based scheme which operates in the Fourier domain for the estimation of translations, rotations and scalings (up to 6 factors) in real face images. They first replaced image functions with complex gray-level edge maps and then compute the standard Cartesian FFT. The next step was to resample the Cartesian FFT on the log-polar grid by using binlinear interpolation. To perform robust correlation, they replace phase correlation with normalised gradient-based correlation schemes (NGC) (Argyriou and Vlachos, 2004). They estimated the rotation and scaling using NGC in the log-polar Fourier domain. Their results showed that their method provides a fast and robust framework for scale-invariant image registration.

Compared with other area-based methods, Fourier based methods are robust to noise and time varying illumination disturbances. Also they have a low computational

complexity and take a fixed period of time in registering any images. In comparison, other area-based methods generally have high computational costs and the time used in registering two images with spatial methods is unpredictable, depending on the images to be registered. The Fourier methods are easy to implement and parallelise. They can be used for generating an initial value for nonlinear optimisation based spatial registration methods.

3.6.1.3 Mutual Information based Methods

In recent times, Mutual Information (MI) has become a popular similarity metric for image registration, especially useful for registering images obtained from different sensors in medical imaging or remote sensing since they are very robust (Zitova and Flusser, 2003, Zibaeifard M. and Rahmati, 2006, Wells III et al., 1996, Maes et al., 1997, Thevenaz P. and Unser M., 2000).

The MI method originates from information theory. MI methods can be considered as a similarity measure, because the maximum value will indicate the best match between a reference image and a sensed image. MI measures redundancy between two images by looking at their intensity distributions and represents a measure of the relative entropy. For two images I_1 and I_2, the definition of the mutual information $MI(I_2, I_2)$ of these two image is as (Gao et al., 2008):

$$MI(I_2, I_2) = H(I_1) + H(I_2) - H(I_1, I_2) \tag{3-23}$$

with $H(I_1)$ and $H(I_2)$ being the entropy of I_1 and I_2, respectively. and $H(I_1, I_2)$ their joint entropy. For image registration, the entropies are calculated by the probability distributions is defined as:

$$H(I_1) = -\sum_{a \in I_1} p_{I_1}(a) \log p_{I_1}(a) \tag{3-24}$$

$$H(I_2) = -\sum_{b \in I_2} p_{I_2}(b) \log p_{I_2}(b) \tag{3-25}$$

$$H(I_1, I_2) = -\sum_{a \in I_1} \sum_{b \in I_2} p_{I_1 I_2}(a,b) \log p_{I_1 I_2}(a,b) \tag{3-26}$$

where a and b are the intensity values in image I_1 and I_2 respectively, $p_{I_1}(a)$ and $p_{I_2}(b)$ are the marginal probability distributions of each image which is the probability of one event, ignoring any information about the other event. $p_{I_1 I_2}(a,b)$ is the joint probability of I_1 and I_2. Therefore, the MI is defined with probability functions as (Cole-Rhodes et al., 2003):

$$MI(I_1, I_2) = \sum_a \sum_b p_{I_1 I_2}(a,b) \cdot \log\left(\frac{p_{I_1 I_2}(a,b)}{(p_{I_1}(a) \cdot p_{I_2}(b))} \right) \tag{3-27}$$

The maximisation of mutual information technique (MMI) was first introduced by Viola and Wells (1997) which has overcome a lot of problems with the registration of multimodal imagery. It has been used in remote sensing by several authors (Chen et al., 2003, Cole-Rhodes et al., 2003).

Cole-Rhodes et al. (2003) have merged mutual information with an optimisation scheme and applied the registration of remotely sensed imagery. By searching for the maximum of the mutual information measure, the best match between two images can be found. To speed up the computation of the similarity measure (MI), they decomposed the input images by a multi-resolution wavelet decomposition which generates images from coarse to fine resolution. Their study is limited to images only

misaligned by rotation and translation. They compared NCC methods with mutual information method by their sharpness at peak values which indicate MI method has a sharper optimum peak than NCC method. However, this is based on the condition that images have been decomposed into multi-resolution levels rather than aligning the original image with template image.

Bao and Ralph (2007) have also examined the use of MI and NCC methods. The experiments are conducted on infrared images with noise. They applied both MI and NCC methods directly on the imagery without using multi-resolution methods. According to their results, they NCC method is more robust that MI method in the presence of noise due to the MI methods have a high sensitivity to errors in the image scaling parameter.

Registration based on mutual information is robust and data-independent and could be used for monomodality and multi-modality images. However, MI methods are like other area-based methods, they require high computation time. In order to speed up the registration process, Zibaeifard and Rahmati (2006) used a multi-resolution approach on medical images which constructs a pyramid of low to high resolution images first and then performs a search from a small number of discrete pixels at coarse level to finer level. An initial value can be obtained by searching low resolution image in a relatively short time. They have improved the method of sample selection which tried to find a suitable subset of image samples depending on the complexity of the regions.

3.6.2 Feature based Methods

Goshtasby et al. (1986) have proposed a sub-pixel accuracy image registration method by extracting closed-boundary regions from digital images. They have discussed registration between day-visual and infrared digital images with

translational, rotational and scaling differences. The centres of gravity of the closed boundary regions are taken as control points. They have designed a method to refine the region boundaries so that optimally similar corresponding regions are obtained. This step enables determination of centres of gravity of corresponding regions. Then, correspondence is established between control points and registration parameters are determined by a least squares error criterion which estimates the transformation parameters by minimizing the sum of squared errors.

Hrka´c and Kalafati´c (2007) presented an approach to multimodal image registration between infrared and visual images taken from the same viewpoint. Their method is based on mapping clouds of points extracted by a corner detector applied to both images. The Harris corner detector is chosen which describes the intensity structure of the local neighborhood for each image pixel (Harris and Stephens, 1988). Then they have adopted two strategies to reduce the false corner points. One is based on the assumption that the images were taken from the same viewpoint so that the corresponding corners in both images should not be too far away. The other is that the point pairs are at least have certain distance away from each other in each image. Hausdorff distance was used to measure the similarity between the images (Vincent, and Laganiere, 2002). Their method uses a small number of parameters which has better results on a wide range of images. However, the corner detection affects the misalignment heavily which needs to be improved by adding more characteristics to corners.

Yang (2001) has proposed an algorithm for SAR and FLIR image registration. The feature based method was chosen due to the poor correlation between FLIR and SAR images which are from different sensors. A two-steps algorithm was used. The initial registration used the available sensor truth information to derive a transformation that maps FLIR images points into the SAR image coordinate. The feature points were extracted by a constant false alarm rate detector. The second stage involved using a generalised Hough transform to complete the registration process. As the generalised

Hough transform is robust to clutter feature points, this algorithm has a certain tolerance to false alarm.

In comparison to area-based methods, feature-based methods are more suitable for situations when illumination changes are expected or multisensory analysis is expected because they work on the extracted features rather than work directly on the image intensity values (Bevilacqua and Azzari, 2007). However, feature-based methods are more sensitive to noise and the features appeared in one image maynot appeared in the others.

3.7 Summary

This chapter has demonstrated some pre-processing methods for FLIR imagery and given a review of existing image registration methods. Due to the limitations of the environmental factors and hardware conditions for infrared imaging systems, the infrared images are very noisy, relatively low contrast and contain dead/saturated pixels which are caused by bad connections between the photo-detectors and the read-out circuits. A contrast enhancement method has been used for improving the intensity distribution of infrared images in section 3.2 and section 3.3 has described the median filter which is used to remove the dead/saturated pixels. To find the pre-processed pixels, a detection algorithm has been proposed in section 3.4. Section 3.5 has presented the basic information of image registration methods including the sources of difference, geometric transformation models, image registration process and interpolations.

Section 3.6 has given a detailed survey of various existing image registration techniques. These included the two main categories: feature-based and area-based. The former are based on extracting the features from images such point, edges, lines, curves or regions. Since the FLIR images have very low contrast and very noisy,

feature-based methods appears less accurate. Area-based methods are generally more reliable than feature-based method due to the fact that they make use of the whole image. Various area-based methods have been investigated.

The area of image registration techniques is vast and can not be explained in full within a single chapter. Moreover, developments in the field are exceedingly rapid and it is almost impossible to keep abreast of all novelties. The presentation of the theoretical background and the algorithm survey sections of this chapter endeavoured to give a comprehensive view of the field what is sought to be solved and how. This chapter has reviews the image registration methods as many aspects as possible, in the best representative way. From each category or subcategory, the most innovative and diverse techniques were described in greater detail. A certain bias was also given towards the description of methods which either relate directly to the contributory part of the thesis or aid its understanding.

The aim of this thesis is to develop an automatic detection system on FLIR images. To detect small targets in infrared imagery successfully, a number of different pre-detection techniques such as image registration, super resolution and image mosaicing techniques will be described in the following two chapters.

Chapter 4 Accurate Registration and Robust Super Resolution Reconstruction

4.1 Introduction

One of the main applications of FLIR systems is in the use of airborne cameras to detect ground objects from imagery of ground features. The effectiveness of FLIR systems is limited by image resolution caused by factors such as the finite size of the infrared focal plane, the optics of the camera and motion blur.

The Super Resolution (SR) technique is a method of increasing the amount of information available to higher level processing algorithms. This technique allows a high resolution image to be reconstructed from a series of low resolution images based on the premise that each image only has sub-pixel movement. By making use of this sub-pixel information a high resolution image can be formed.

The primary factor that controls the quality of the super-resolved image is the accuracy of registration of the low resolution frames. In the previous chapter an full review has been given on image registration techniques. Registration of FLIR imagery is more challenging than that of visible band images due to the fact that FLIR images have extremely low signal to noise ratios (SNR) and can only provide limited information for performing image registration. Furthermore, FLIR images are more easily cluttered by non-target objects (such as background noise) and the brightness can also be suppressed by the clouds due to flying above clouds. An image registration technique for FLIR images needs to consider all of these issues.

This chapter will present an automatic sub-pixel image registration method for low contrast FLIR imagery in the presence of noise in the section 4.2. Following this, in Section 4.3 an effective and robust super resolution algorithm has been proposed for FLIR images. Finally, a summary will be given in Section 4.4. An overview of the super resolution image reconstruction algorithm is shown as Figure 4.1.

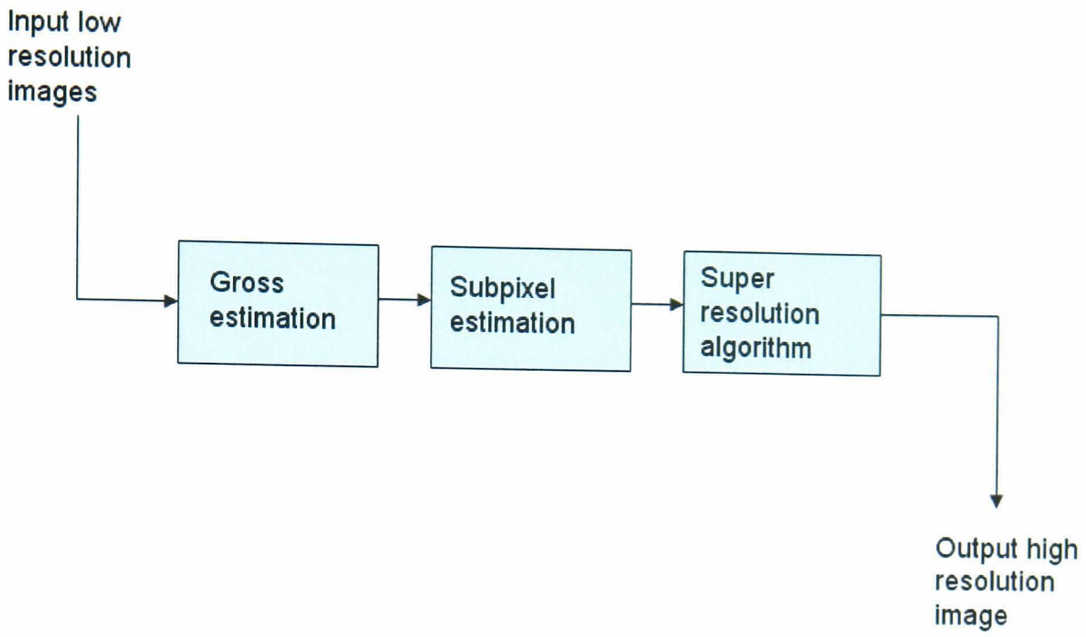


Figure 4.1 An overview of the super resolution image reconstruction algorithm

4.2 Accurate Image Registration on FLIR Imagery

The accuracy of image registration has a direct influence on the effectiveness of super resolution algorithms and requires sub-pixel accuracy. In selecting an appropriate registration algorithm, there are certain factors that need to be considered: the type of transformation between images, whether the images have salient features, and what distortions are present in the images.

As real FLIR imagery has very low contrast and is contaminated by a large amount of noise, feature-based registration methods are generally inappropriate because very few features are presented constantly in a sequence of images. Therefore in this project, feature-based registration algorithms have been excluded from the candidate registration algorithms for FLIR imagery and area-based methods are mainly considered here.

In FLIR imagery, image movement is mainly induced by camera movement and vibration of the moving platform. As the imagery is generally taken at certain altitude, a small change in height can be ignored with respect to the total. Therefore, it is generally a rigid transformation without any significant scale change for straight downward rather than forward or backward looking cameras.

The difficulties in aligning FLIR imagery can be caused by various sources. There are situations when cloud occludes the background in the images or when the images are taken at different times and temperature differences introduce variation between frame intensities.

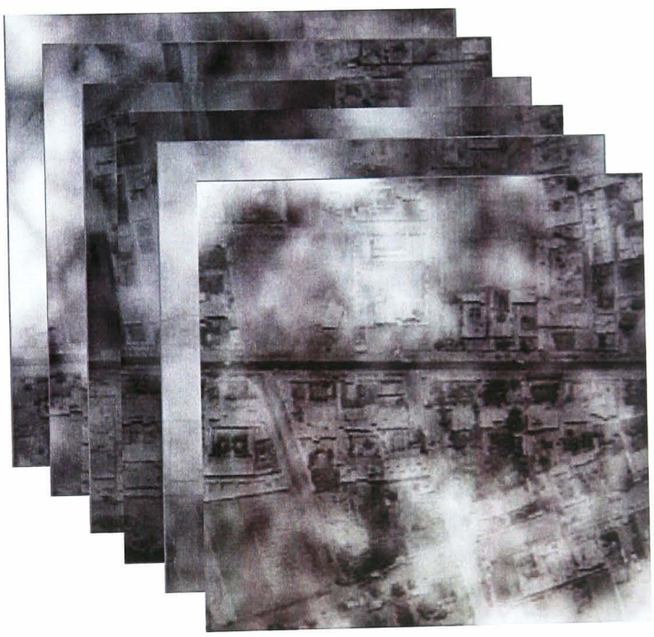
As well as image motion caused by camera movement, moving objects in the scene also introduce motion factors. These types of object are described as Independently Moving Objects (IMOs) (Strehl and Aggarwal, 2000). In the context of airborne imagery, moving ground objects appear relatively small and consequently the background of the scene will cover most of the image. This means that the dominant motion will be the displacement of the background caused by the observer motion (ego-motion). In this context, independently moving objects can be understood to be objects whose motion violates the dominant motion model. Image registration techniques can be used to remove the dominant ego-motion movement and leave the IMOs which can be detected by later processing.

Most image registration methods are based on a frame-to-reference model, which means each frame is aligned to the same reference image. This is generally not suitable in FLIR applications because if the sequences are very long, latter frames do not have enough overlapped area or do not overlap at all with the reference image. Therefore, a suitable image registration system for FLIR imagery needs to be designed so as to register images without reference to a fixed template.

The most optimum technique for FLIR image registration is one which utilises the entire image, can handle large displacements and is robust against object motion and global alignment.

In this research, an automatic image registration method has been developed. This is a specific implementation of the general registration techniques discussed in the previous chapter. This technique uses two stages of algorithm; the initial stage uses Fourier registration to align images with large translations and rotations. Fourier registration is computationally efficient and robust to noise. Subsequently, a normalised cross correlation registration method with a stochastic gradient search is applied to adjust the image to sub-pixel accuracy. This automatic image registration method has been tested both on synthetic and real images with various backgrounds and occlusions. Examples of FLIR image sequences with and without cloud are shown in Figure 4.2. The generation of synthetic images has been described in chapter 2.

(a)



(b)

Figure 4.2 Examples of synthetic FLIR image sequences: (a) Kandarhar image sequences without cloud occlusion, (b) Kandarhar image sequences with synthetic cloud occlusion.

4.2.1 Gross Estimation Method

4.2.1.1 Pre-filtering

Since FLIR images are often contaminated by noise, pre-processing is necessary to reduce its effect. Since noise contributes high spatial frequencies in an image it can be removed by use of a smoothing filter. Gaussian filters are a class of linear smoothing filters with weights chosen according to the shape of a Gaussian function. Gaussian smoothing reduces higher spatial frequencies more than the lower spatial frequencies and is a very good filter for removing noise drawn from a normal distribution. The discrete Gaussian function in two dimensions is:

$$\begin{aligned} G(x) &= \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2}{\sigma^2}\right\}\times\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{y^2}{\sigma^2}\right\} \\ &= G(x)\times G(y) \end{aligned} \tag{4-1}$$

where the Gaussian spread parameter σ determines the width.

A large σ implies a wider Gaussian filter and greater smoothing. There are several reasons to choose a Gaussian filter as the pre-filter. Firstly, the Gaussian filter is rotationally symmetric, which means that the amount of smoothing performed by the filter will be the same in all directions. Second, the smoothed image by the Gaussian filter will not be corrupted by contributions from unwanted high-frequency signals, while most desirable signals will be retained. In addition, large a Gaussian filter can be implemented very efficiently since a two dimensional Gaussian can be operated by convolving the image with a one-dimensional Gaussian and then convolving the result with the same one-dimensional filter oriented orthogonally to the Gaussian used in the first stage.

The Fourier transform assumes that an image is cyclic, which means that the first image row follows the last image row and the first image column is a continuation of the last image column. Therefore, if image intensities near the top and bottom or near the left and right image borders are different, direct computation of image smoothing using may be more accurate and faster than computation by the FFT algorithm (Goshtasby, 2005).

4.2.1.2 Fourier Method

The 2D Fourier transform for a digital image f can be defined as (Arfken, G., 1985, Sonk et. al, 2008):

$$F(u,v)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\exp\left[-2\pi j\left(\frac{xu}{M}+\frac{yv}{N}\right)\right], \qquad (4\text{-}2)$$

$$u=0,1,\cdots,M-1,\ v=0,1,\cdots,N-1.\omega=2\pi j$$

The Fourier transform of an image $f(x,y)$ is a complex function and each function has a real part $R(\omega_x,\omega_y)$ and an imaginary part $I(\omega_x,\omega_y)$ at each frequency (ω_x,ω_y) of the frequency spectrum:

$$\begin{aligned}F(\omega_x,\omega_y)&=R(\omega_x,\omega_y)+jI(\omega_x,\omega_y)\\&=|F(\omega_x,\omega_y)|\exp[j\phi(\omega_x,\omega_y)],\ j=\sqrt{-1}\end{aligned} \qquad (4\text{-}3)$$

where $|F(\omega_x,\omega_y)|$ is the magnitude or amplitude of the Fourier transform and where $|\phi(\omega_x,\omega_y)|$ is the phase angle.

Fourier registration estimates the motion parameters between the template image and

other images in the frequency domain. Motion can be described as a function of three parameters: horizontal and vertical shifts, t_x and t_y, and a planar rotation angle ϕ. Assume $I_1(x,y)$ and its shifted and rotated version $I_2(x,y)$:

$$I_2(x,y) = I_1\left(R\left((x,y)+(t_x,t_y)\right)\right), \tag{4-4}$$

with rotation matrix $R = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$.

A frequency domain approach enables estimation of the horizontal and vertical shifts, and the planar rotation separately.

4.2.1.2.1 Translation Registration

The translation registration is based on the shift theorem of the Fourier transform. Given two images I_1 and I_2 which differ only by a displacement (t_x,t_y):

$$I_2(x,y) = I_1(x-t_x, y-t_y), \tag{4-5}$$

Their corresponding Fourier transforms F_1 and F_2 is:

$$F_2(\omega_x,\omega_y) = \exp\{-j(\omega_x t_x + \omega_y t_y)\}F_1(\omega_x,\omega_y). \tag{4-6}$$

Therefore, the phase difference is directly related to their displacement which is given by: $\exp\{-j(\omega_x t_x + \omega_y t_y)\}$.

The cross-power spectrum of two images is defined as:

$$\frac{F_1(\omega_x,\omega_y)F_2^*(\omega_x,\omega_y)}{\left|F_1(\omega_x,\omega_y)F_2^*(\omega_x,\omega_y)\right|} = \exp j(\omega_x t_x + \omega_y t_y) \tag{4-7}$$

where F^* is the complex conjugate of F, the phase of cross-power spectrum is equivalent to the phase difference between images.

The Fourier method searches for the optimal match according to information in the frequency domain. A flowchart of the local Fourier registration method for translational registration is shown in Figure 4.3. To register the input image with the template image, a Gaussian filter has first been applied to both. Then, after both images have been transformed into the frequency domain, the cross-power spectrum between the images can be calculated. By searching for the maximum cross spectrum value after inverse transform, the shift between images can be found and the transformation matrix parameters can be determined. There are two functions in the Matlab image processing tool box that can be used to transform the images. The function `maketform` is used to create geometric transformation structures and `imtransform` is used to apply the 2D spatial transformation to the input image.

Experiments using the Fourier registration method have been conducted on synthetic imagery with and without cloud. The shift registration results are shown in Figure 4.4 and 4.5.

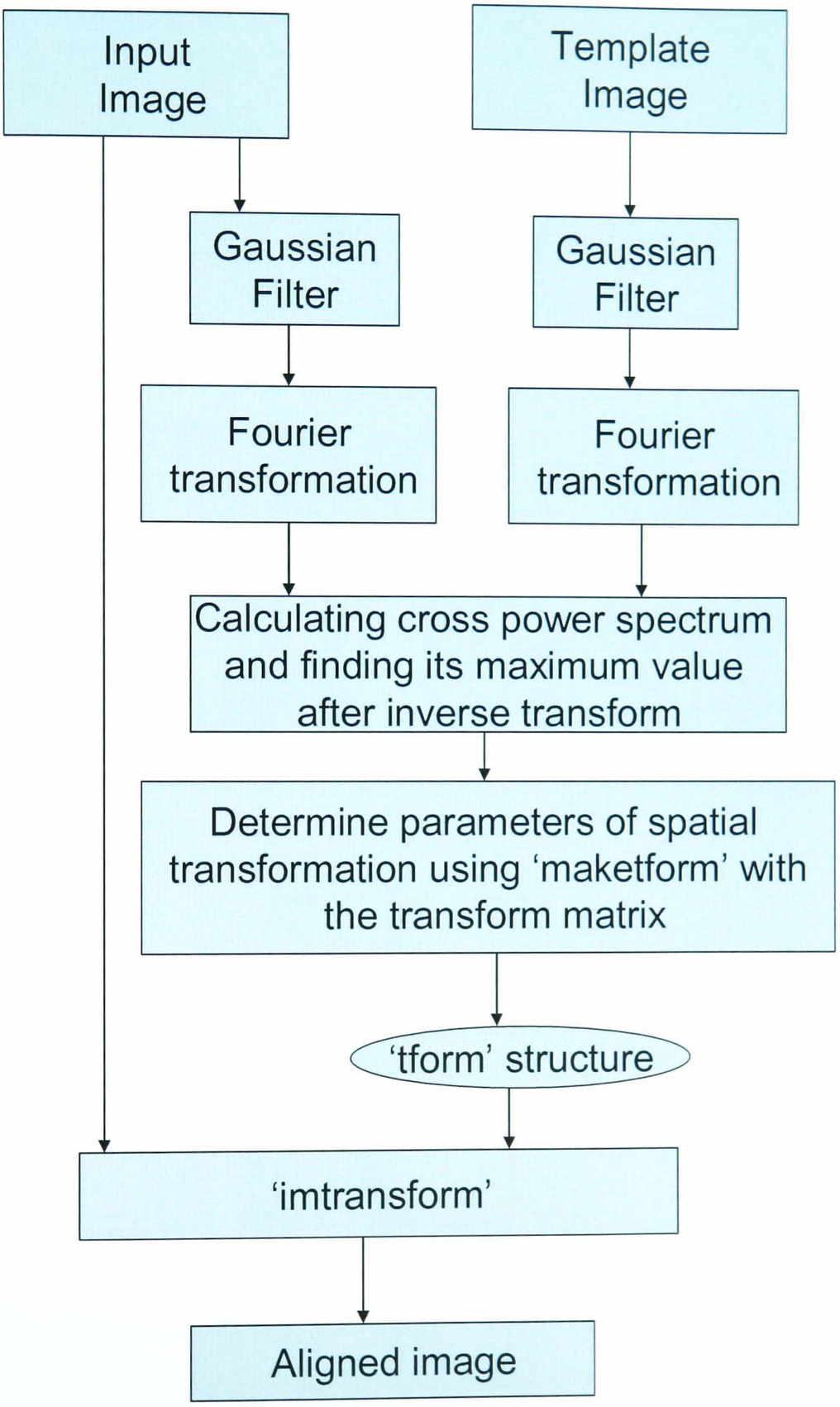
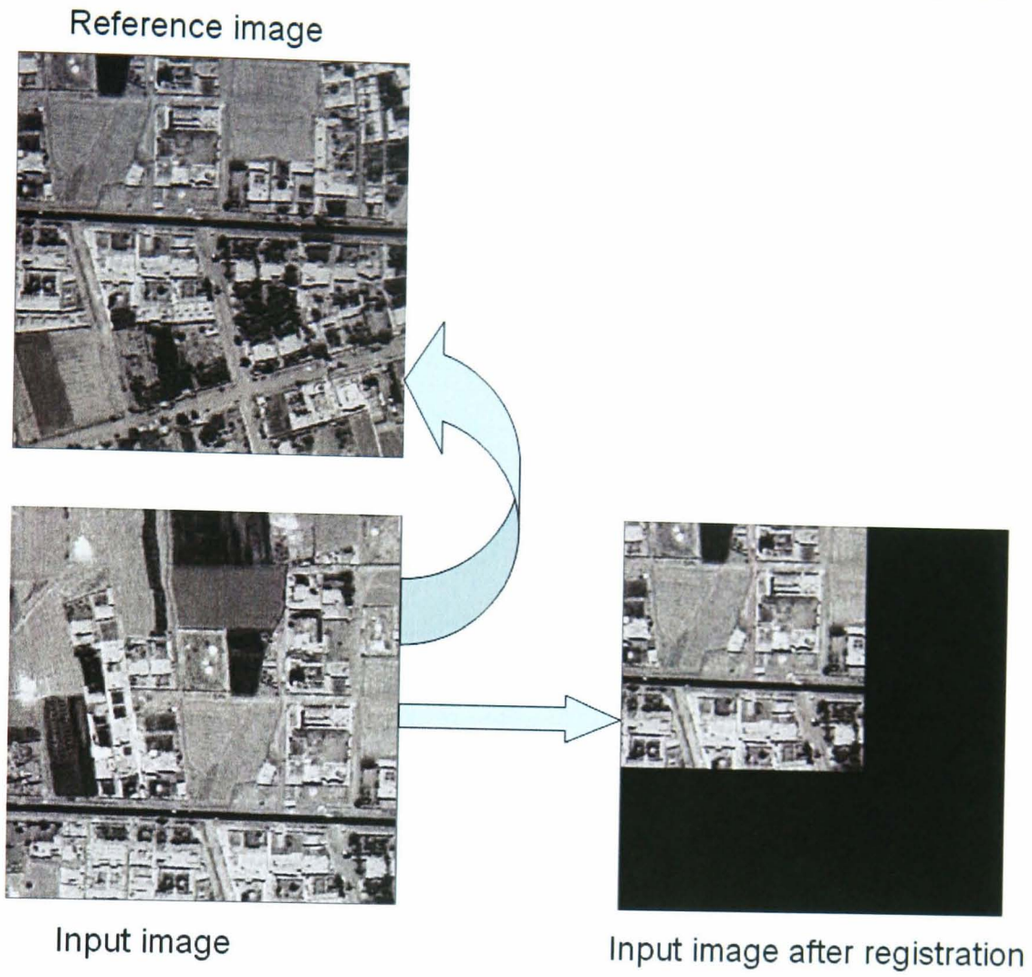


Figure 4.3 Flowchart of local Fourier registration process for translations



(a)

(b)

Figure 4.4 Example of shift registration images: (a) synthetic image registration without cloud (b) registered image overlapped with the reference image.

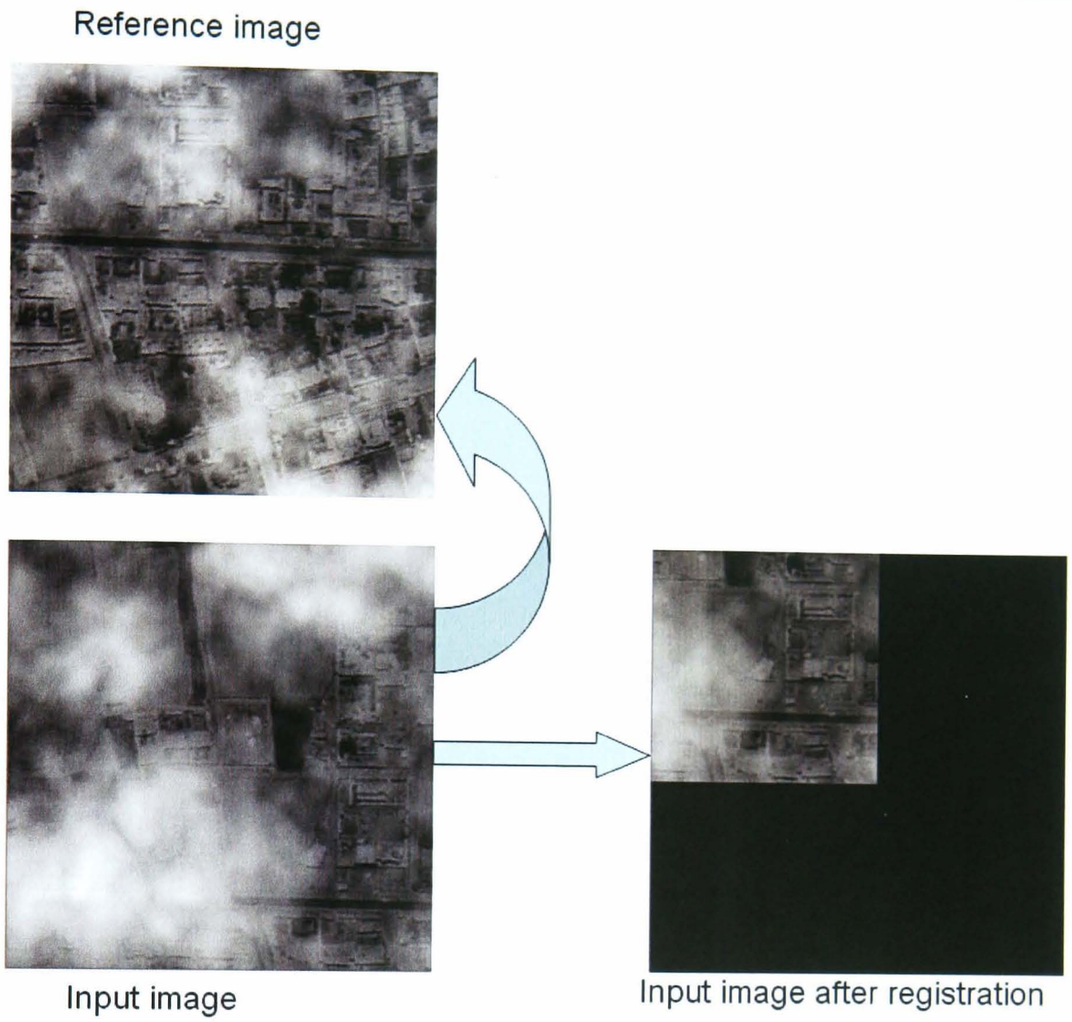


Figure 4.5 Image shift registration on image sequences with cloud cover

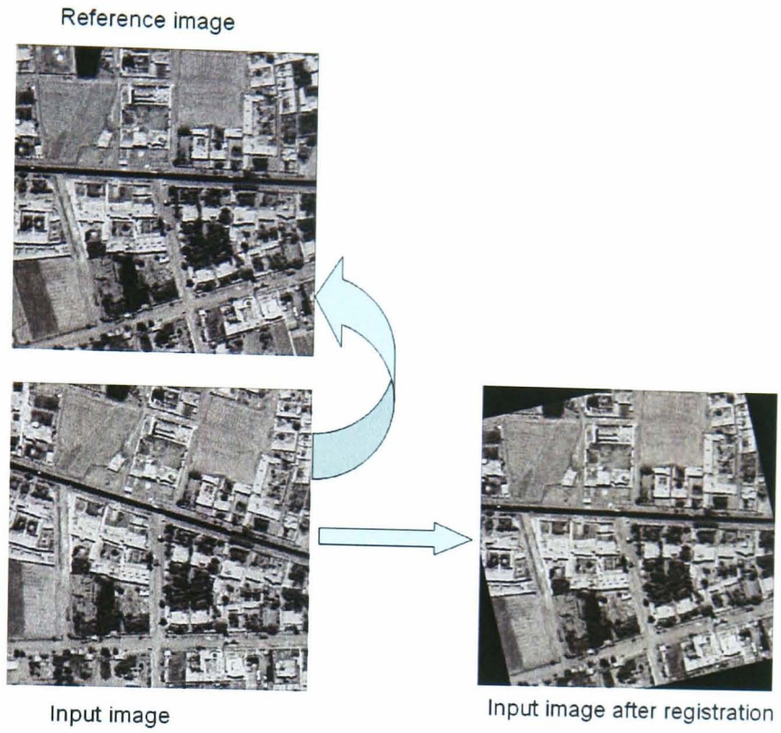
4.2.1.2.2 Rotation Registration

Rotation registration is based on the rotation invariant theory. When there are pure translations present between two images, phase correlation has a maximum peak, and the corresponding location gives the translation parameters (t_x, t_y) (Vandewalle, et al., 2006).

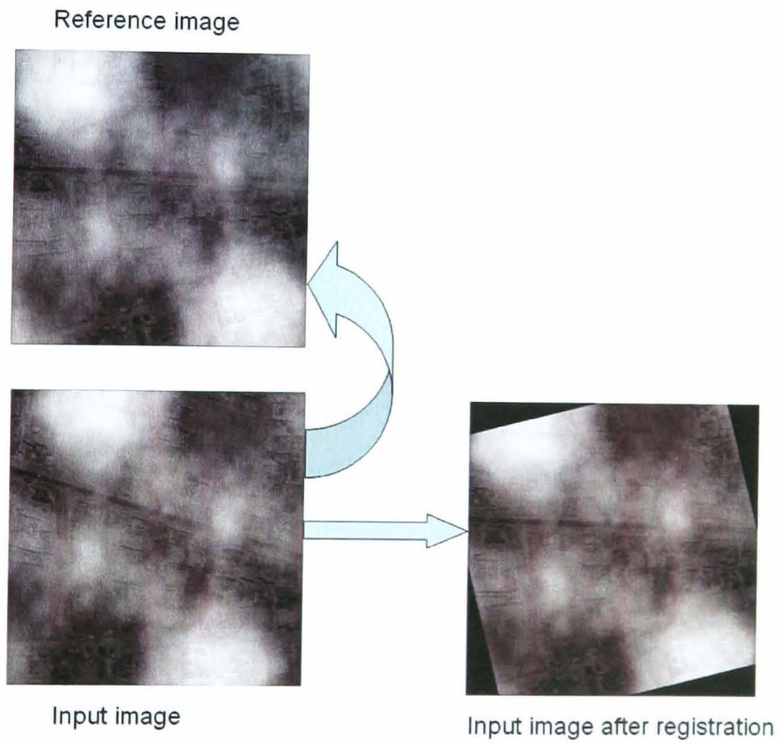
Suppose the two images $I_1(x,y)$ and $I_2(x,y)$ to be registered involve both translation and rotation, with the angle of rotation being ϕ between them, their transforms are related by:

$$I_2(x,y) = \exp\{-j2\pi(xt_x + yt_y)\}I_1(x\cos\phi + y\sin\phi, -x\sin\phi + y\cos\phi) \quad (4\text{-}8)$$

When $I_2(x,y)$ is rotated by ϕ, there will be only translation left between the images, and the phase correlation with $I_1(x,y)$ should give a maximum peak. So by rotating $I_2(x,y)$ by 0.01 degree each time and computing the correlation peak for that angle, when there is a highest peak value, that angle is the rotation angle. The range of rotation is set between -10 degrees to +10 degrees to reduce the computational complexity, and is based on the fact that the camera on the moving platform does not rotate more than this range. The experimental results on imagery with rotational movement are shown in Figure 4.6.



(a)



(b)

Figure 4.6 Examples of rotation registration between frames: (a) image sequences without cloud, (b) image sequences with cloud

4.2.1.2 Global Image Registration

Generally, the image alignment requires a consistent appearance between two images. When registering a sequence of images, the registration between two successive frames can be considered as local registration and the registration of the whole sequences to the same coordinate system can be considered as global registration.

Among a sequence of images, one image can be chosen as the template image and all the other images can be aligned with this template image if they have some overlapped area. However, for the alignment of long image sequences, when there is no overlapping area between the template image and some of the other frames, or the overlapping area is less than 50% of the image content, it is impossible to select a fixed template image for all the other frames because not all of them have spatial overlap with the template image. A solution for this dilemma is to find global transformations of the image sequences.

Global registration approaches align groups of frames by considering both their spatial and temporal contiguity. Global alignment is generally achieved by finding all the transformations from local alignments of adjacent images, then converting the transformations into a global transformation. An illustration of global registration is shown in Figure 4.7, where T is the transformation matrix between the first and last image and is obtained by combining all five transformations between adjacent frames (Winkelman and Patras, 2004).

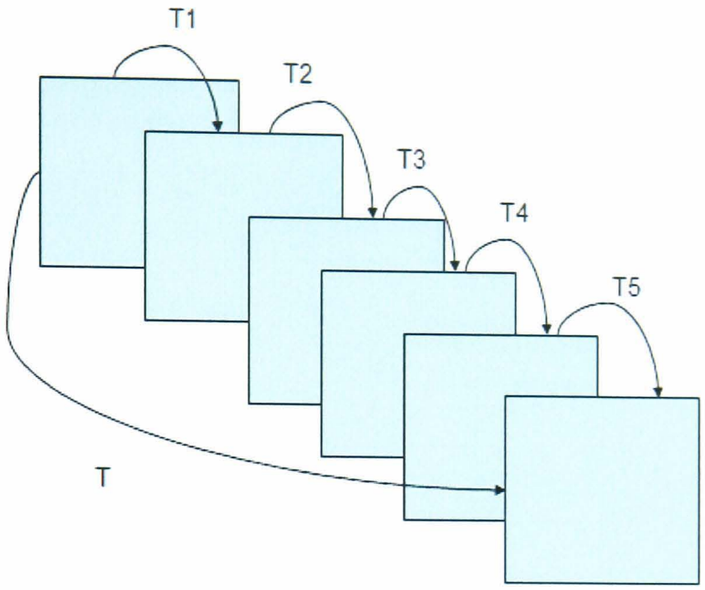


Figure 4.7 An illustration of global registration

The transformation function of each local pair of images is adjusted to a global registration by combining the present transformation with the global transformation. A flowchart of a global Fourier registration process is shown in Figure 4.8. The registration has been conducted between neighboured frames, and each registration generates a Matlab structure `tform` by `maketform`. These local `tform` structures can be combined to generate a global `tform` structure, which can be used for registering the input images with the template image. By combining the transforms, the error in registration tends to accumulate, whereas registering back to the first frames tends to generate a smaller error- it is a conflict between having a larger common overlap area to do the registration and reducing the accumulation error.

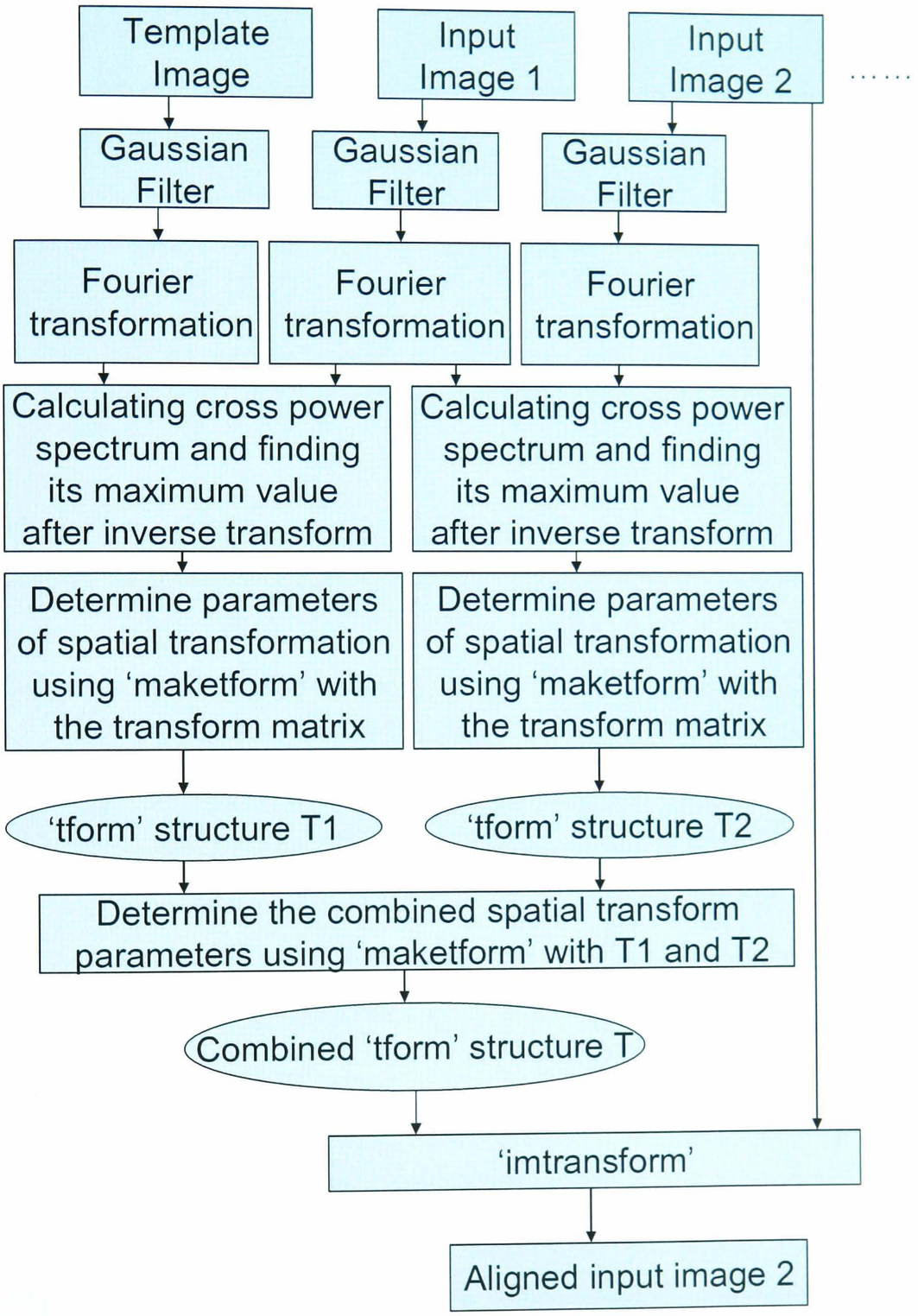


Figure 4.8 Flowchart of global Fourier registration process

Global registration can reduce the error that occurs during the process of registration with the same template image. Figure 4.9 shows an example of image shift registration error on image sequences with cloud cover by aligning to the same reference image.

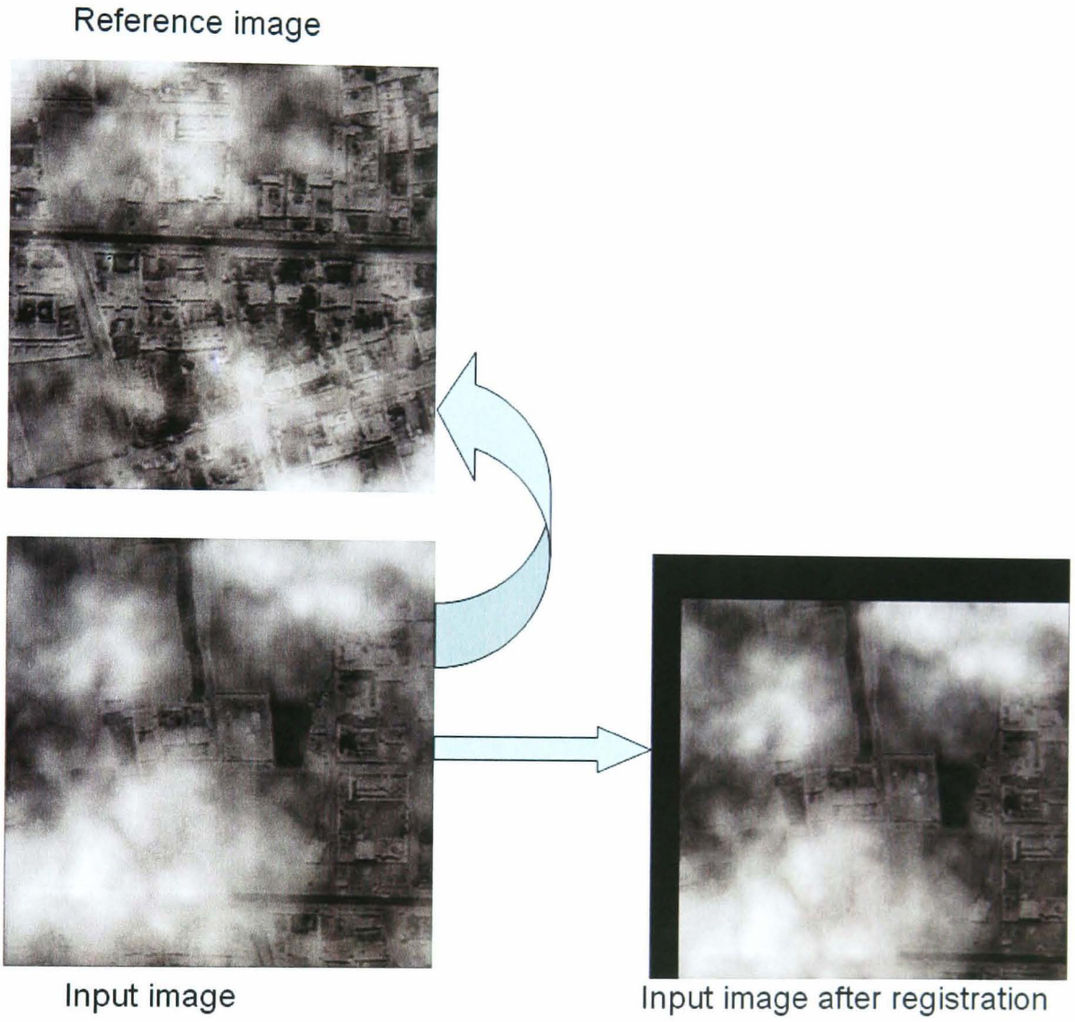


Figure 4.9 An example of image shift registration error on image sequences with cloud cover by aligning to the same template image

Figure 4.10 shows the registered shifts by global Fourier registration method and the same template Fourier registration method. Figure 4.10 (a) is the result for image sequences without clouds by the same template Fourier registration between the input image and the template image; Figure 4.10 (b) is the result for image sequences without clouds by global registration; Figure 4.10 (c) is the result for image sequences with clouds by the same template Fourier registration; while Figure 4.10

(d) is the result for image sequences with clouds by global registration. By comparing each graph in Figure 4.10, it is clear that same template Fourier registration method generates more errors in the situation with cloud cover.

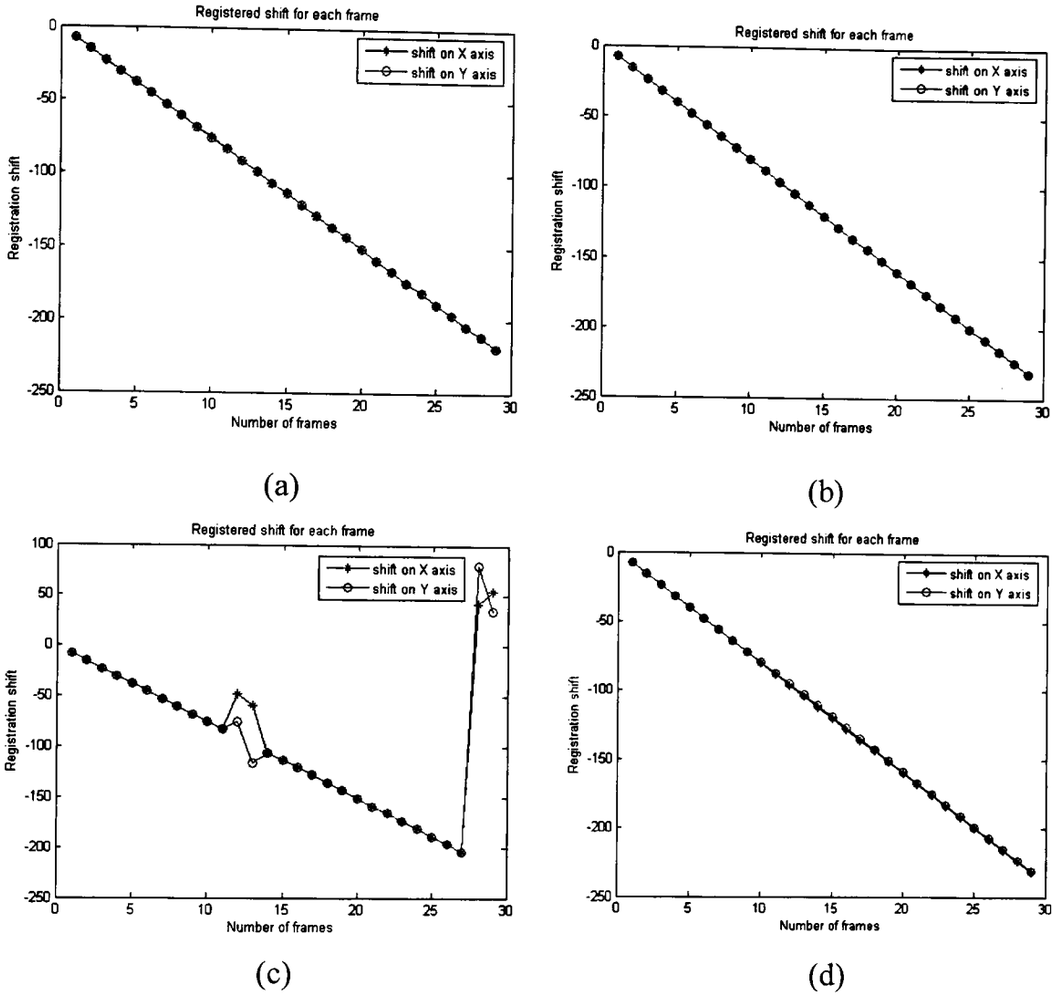


Figure 4.10 Registered shifts between each frame and first frame by Fourier registration method: (a) result for image sequences without clouds by the same template Fourier registration, (b) result for image sequences without clouds by global registration, (c) result for image sequences with clouds by the same template Fourier registration, (d) result for image sequences with clouds by global registration.

So as to test the accuracy of the image registration methods, Fourier translations registration has been tested on 53 sets (30 frames each) of Kandarhar image sequences, with and without cloud. By comparing known data with the transformation data obtained by registration, the average error of registration can be

obtained, as shown in Figure 4.11. Either for image without or with cloud, the global Fourier registration has shown its accuracy and robust since there is no error over pixel size according to the results. However, for the same template registration method, there are large scale errors especially after frame 16 due to the fact that there is not enough overlapped area between the current frame and template frame.

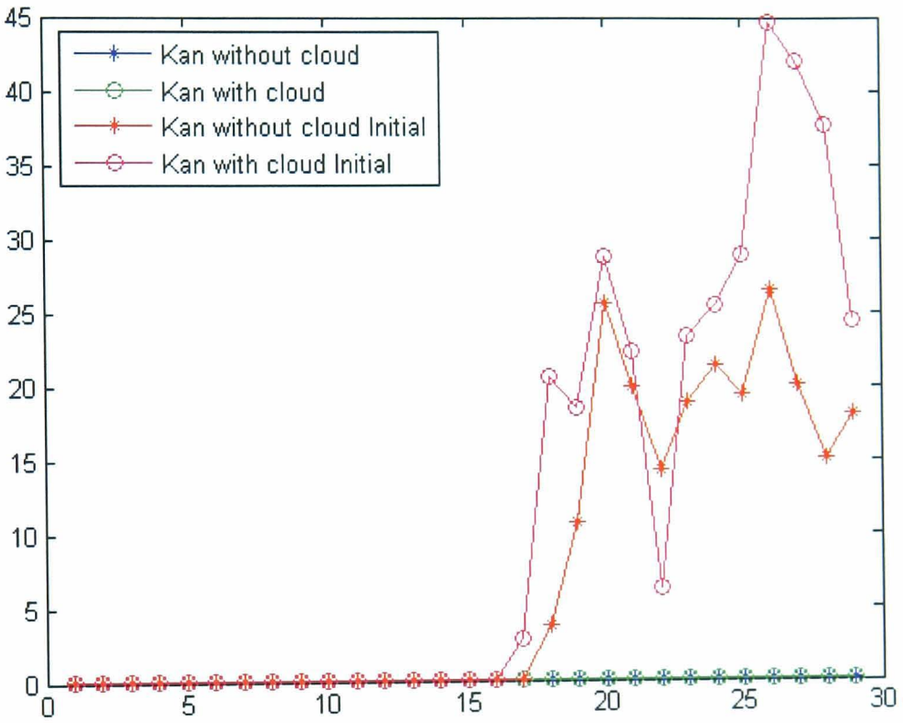


Figure 4.11 Average error of Fourier registration on Kandarhar images

Figure 4.12 and Figure 4.13 show the experimental results of global Fourier registration on synthetic translated and rotated images with and without cloud cover. Both sets of experimental results show a good performance of global Fourier registration method.

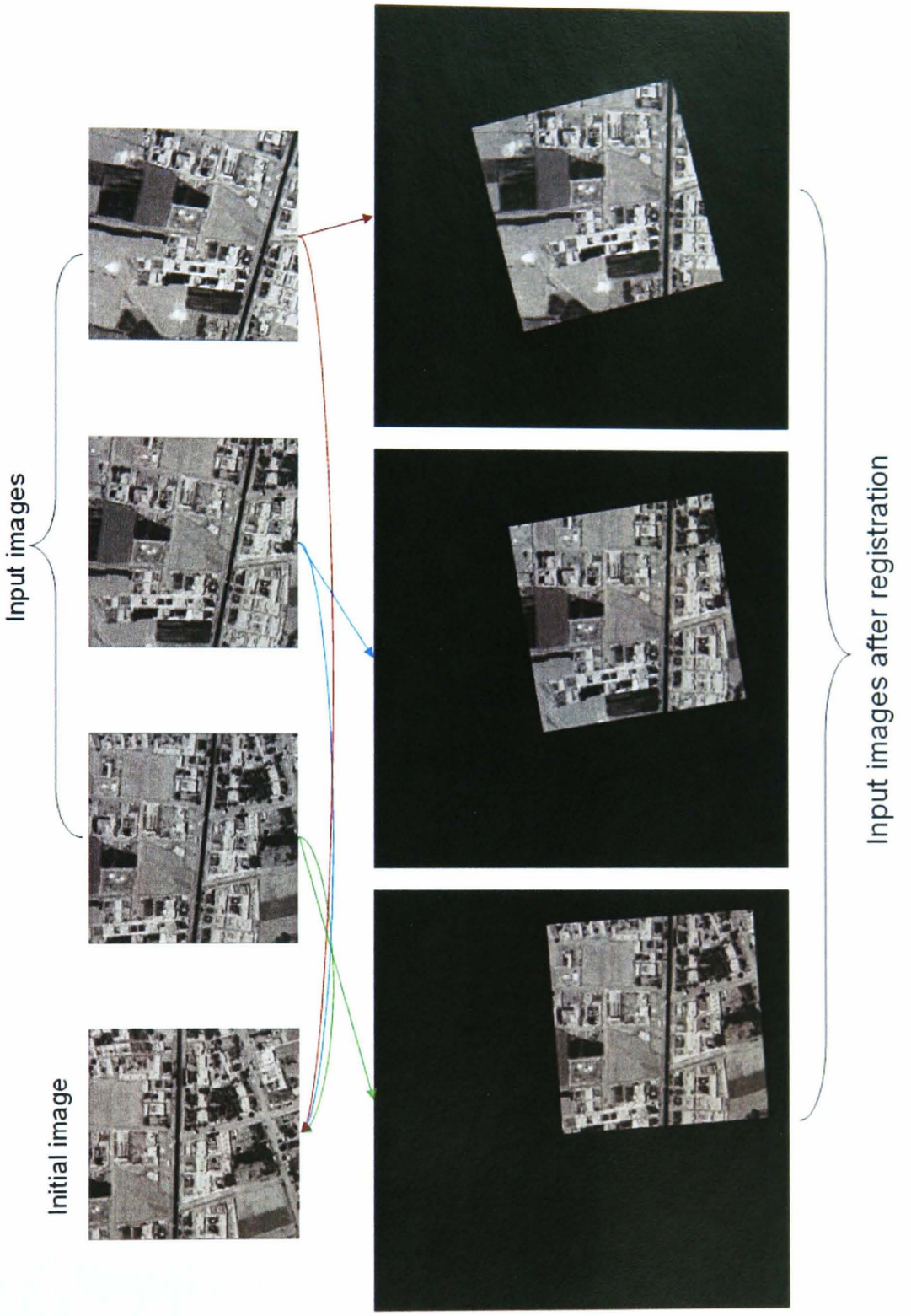


Figure 4.12 An example of image registration on FLIR images with rotation and translation

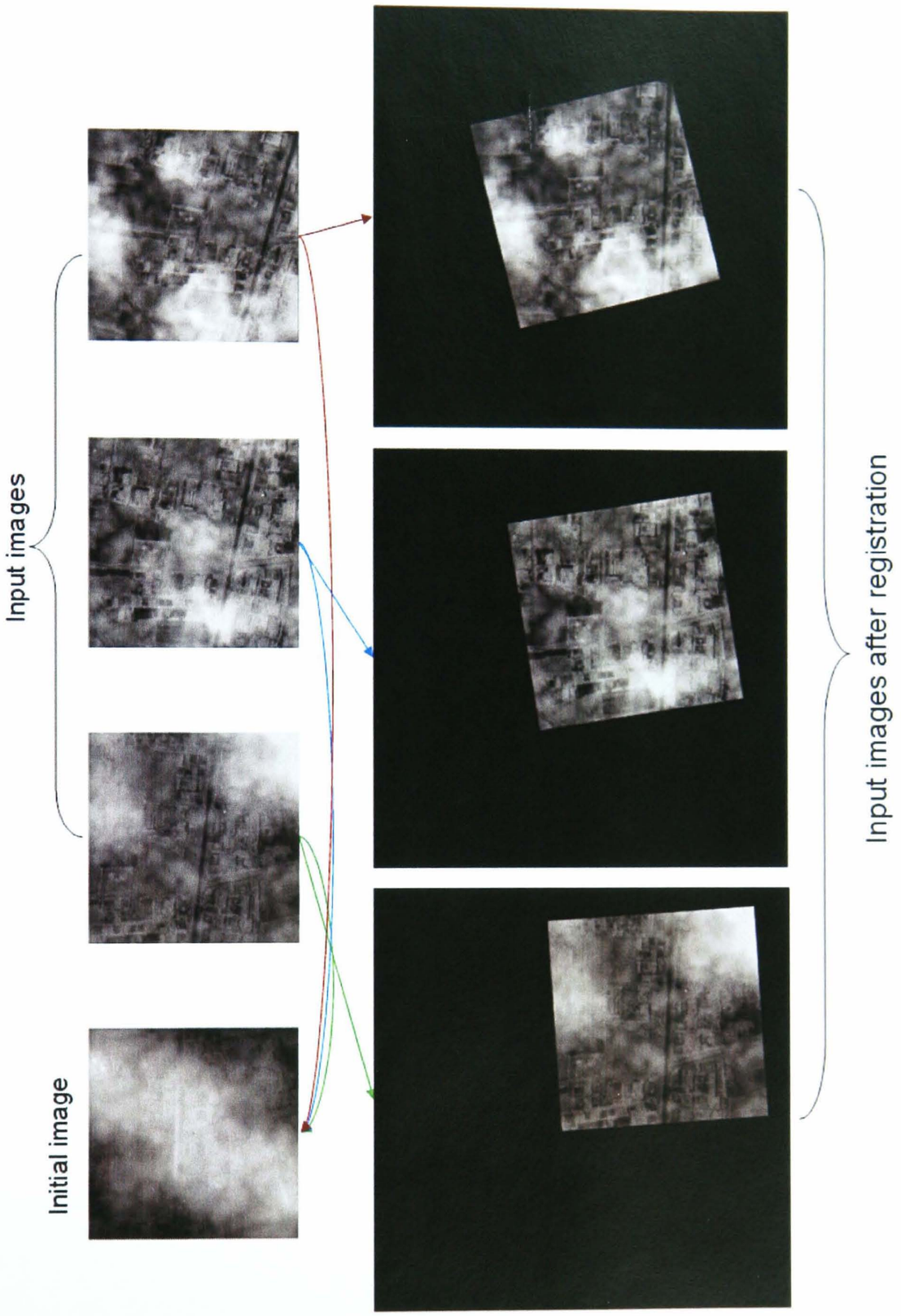


Figure 4.13 An example of mage registration with rotation and translations in presence of clouds

4.2.2 Sub-Pixel Estimation Methods

The sub-pixel registration estimation is achieved by the combination of Fourier method and cross correlation method. A flow chart of proposed method is shown in Figure 4.14. The next step in the registration task is to eliminate the residual registration error following Fourier registration. A normalised cross correlation method combined with a stochastic gradient search has been adopted to reach sub-pixel accuracy.

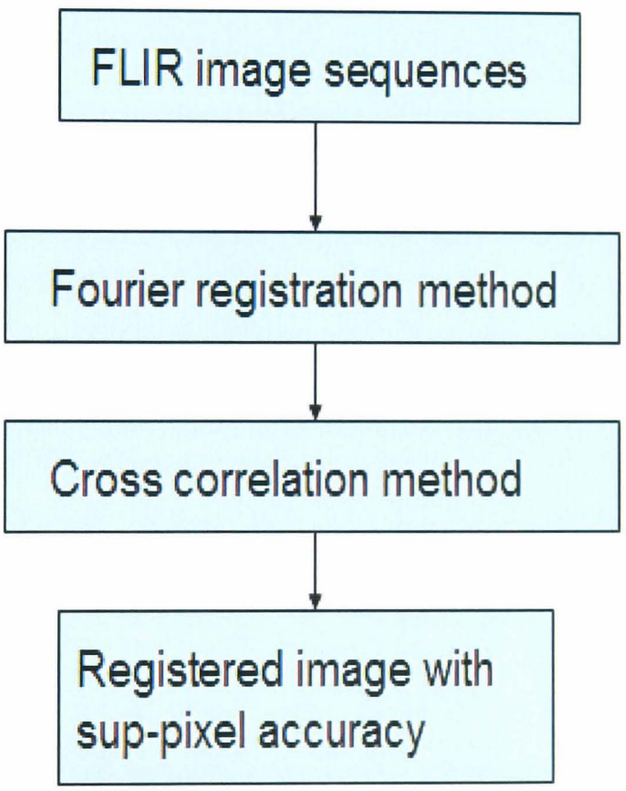


Figure 4.14 A flow chart of sub-pixel estimation method

4.2.2.1 Normalised Cross Correlation Method

For a template image I_1 and sensed image I_2, the two dimensional normalised cross correlation function measures the similarity for each translation (Brown, 1992):

$$C(t_x,t_y)=\frac{\sum_x\sum_y I_1(x,y)I_2(x-t_x,y-t_y)}{\sqrt{\left[\sum_x\sum_y {I_2}^2(x-t_x,y-t_y)\right]}}\cdot\frac{\mathrm{cov}(I_1,I_2)}{\sigma_{I_1}\sigma_{I_2}}$$

$$=\frac{\sum_x\sum_y\left(I_1(x,y)-\mu_{I_1}\right)\left(I_2(x-t_x,y-t_y)-\mu_{I_2}\right)}{\sqrt{\sum_x\sum_y\left(I_2(x-t_x,y-t_y)-\mu_{I_2}\right)^2\sum_x\sum_y\left(I_1(x,y)-\mu_{I_1}\right)^2}} \qquad (4\text{-}9)$$

If the template matches the image exactly, except for an intensity scale factor, at a translation of (x,y), the cross correlation will have its peak at $C(t_x,t_y)$. Therefore, it is possible to find the degree of similarity between the images by computing C over all possible translations.

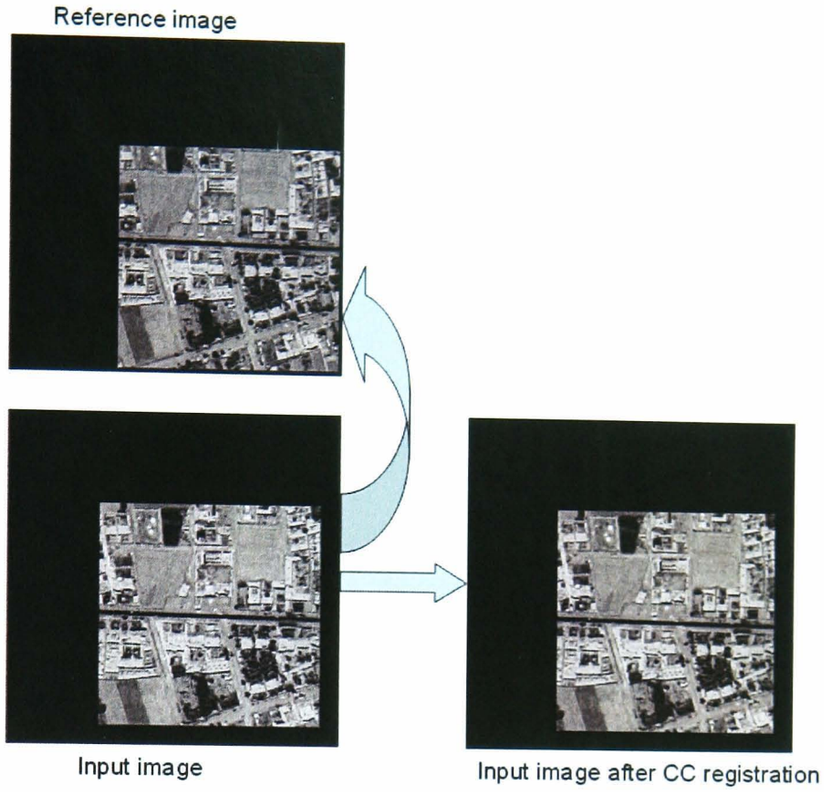
4.2.2.2 Stochastic Search Method

To find the transformation that maximises the value of the cross correlation, a stochastic gradient search method is employed here. This is because after the Fourier method has been applied to correct for gross shifts, the residual movement of images is relatively small. This search method uses a set of arbitrary control points to define a two-dimensional affine or projective transformation. The control points are chosen to be distributed around the images to be registered – to give a large baseline for the geometric transformation to be found – but they do not correspond to features within the image itself.

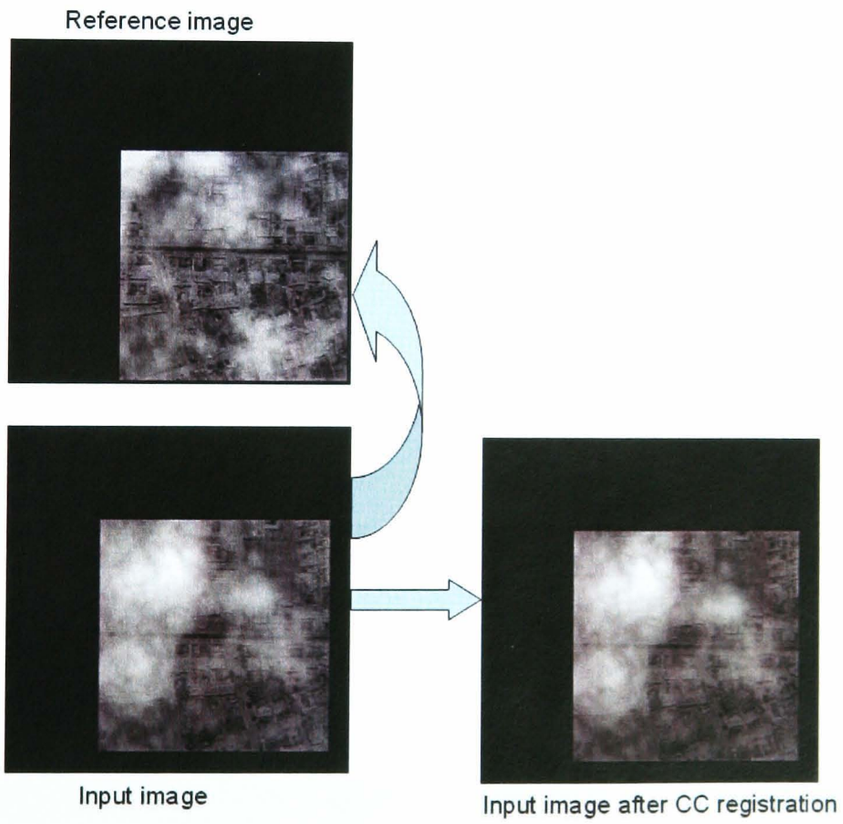
The number of control points determines the dimensionality of the search space: three points for a projective transform give a six-dimensional search space. The stochastic search algorithm selects a random direction in the search space (selecting random perturbations to the *x*/*y* co-ordinates of the control points) and calculates the gradient of the cross-correlation function along this direction. The algorithm then

steps in the direction where the gradient is positive, hopefully towards a global maximum. The size of the step is proportional to the estimate of the gradient, thereby increasing the rate of convergence to the maximum. In this case, the fact that the gross shift required for image registration has already been found using the Fourier transform method, means that the convergence to a local (rather than global) maximum is unlikely, and in practice turns out to be the case for the scenarios studied here.

The experimental results of cross correlation registration are shown as Figure 4.15. By comparing the Figure 4.15 (a) and Figure 4.15(b), the cloud cover does not have too much effect on the cross correlation method.



(a)



(b)

Figure 4.15 Examples of cross correlation registration: (a) registration between frames without cloud (b) registration between frames with cloud

4.3 Super Resolution Techniques

4.3.1 Basic Theory

High resolution images are often required in applications such as remote sensing, military and medical imaging because they provide additional detail that may be critical in accurately analysing the information.

Current CCD technology has almost reached its limit in achieving high resolution, and the cost of high resolution cameras is relatively high (Chaudhuri, 2002). A potentially good alternative is to enhance the spatial resolution of images using signal processing techniques to obtain a high resolution (HR) image (or sequence) from multiple low resolution (LR) images (Park et al., 2003). An illustration of super resolution reconstruction from low resolution sequences is shown in Figure 4.16.

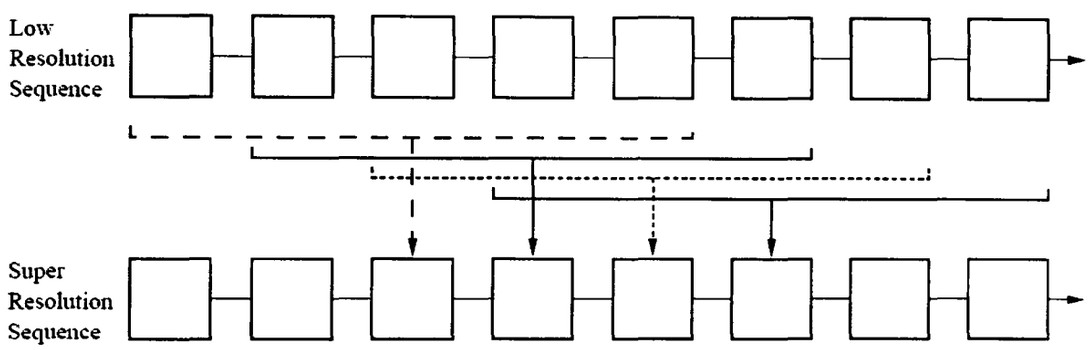


Figure 4.16 Super-resolution reconstructions from a low resolution image sequence (Borman and Stevenson, 1998).

Image super-resolution is the process of reconstructing a high-resolution image from a single image or a sequence of noisy, low-resolution image frames of the same scene but with sub-pixel shifts. If the images were only shifted by integer amounts, then each image would contain the same intensity values at the same spatial locations and consequently there would be no new information that can be used. In this case a simple interpolation scheme (bilinear, cubic spline, etc.) can be used to increase

resolution. Interpolation methods are normally used to increase the size of a single image (Schoenberg, 1969, Unser et al., 1995). Interpolation method can be used to find an interpolated value between pixels in one image. However, the values that obtained from the up-sampled image tend to blur edges because they try to impose a continuous transition model onto regions of different intensity. It is the same when using just average over a number of images and the image tend to be blurry.

If the images have sub-pixel shifts, each image contains different information and the new information contained in each low-resolution image can be exploited to obtain a high-resolution image.

4.3.1.1 Mathematical Formula for Super Resolution

The super resolution reconstruction problem can be analysed by formulating an observation model that relates the original high resolution image to the observed low resolution images. Consider the desired high resolution image of size $L_1N_1 \times L_2N_2$ written in lexicographical notation as the vector $x = [x_1, x_2, \cdots, x_N]^T$, where $N = L_1N_1 \times L_2N_2$. x is the ideal un-degraded image that is sampled at or above the Nyquist rate from a continuous scene, which is assumed to be band-limited.

The size of each low resolution image can be considered $N_1 \times N_2$ with L_1 and L_2 representing the down-sampling factors in the observation model for the horizontal and vertical directions respectively. The kth low resolution image is denoted in lexicographic notation as $y_k = [y_{k,1}, y_{k,2}, \cdots, y_{k,m}]^T$, for $k = 1,2,\cdots,p$ and $M = N_1 \times N_2$.

The high resolution image x is considered to remain constant during the acquisition

of the multiple low resolution images, except for any motion and degradation allowed by the model. Therefore, the observed low resolution images result from, warping, blurring and sub-sampling operators performed on the high resolution image x and each low resolution image is corrupted by additive noise. A block diagram for the observation model is illustrated in Figure 4.17.

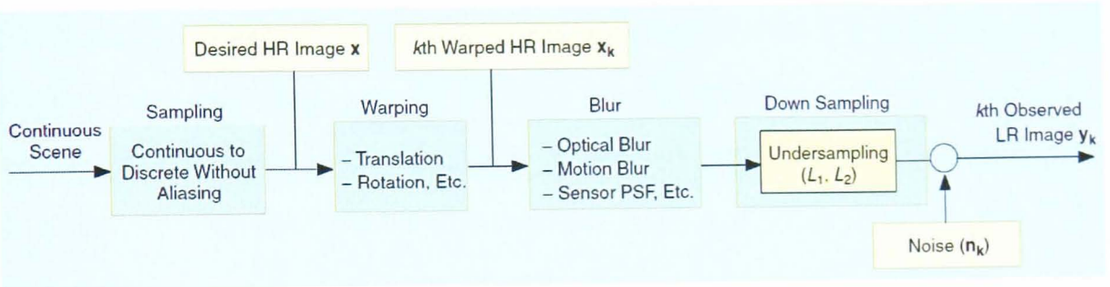


Figure 4.17 Observation model relating LR images to HR images (Park et al., 2003)

The observation model can be expressed mathematically as (Elad and Feuer, 1997, Rajan and Chaudhuri, 2001):

$$y_k = DB_k M_k x + n_k \text{ for } 1 \le k \le p \tag{4-10}$$

where M_k is a warp matrix of size $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, B_k represents a $L_1N_1L_2N_2 \times L_1N_1L_2N_2$ blur matrix, D is a $(N_1N_2)^2 \times L_1N_1L_2N_2$ sub-sampling matrix, and n_k represents a lexicographically ordered noise vector.

The warping matrix M_k may contain global or local translations, rotations etc. The warping factor can be registered by the use of registration techniques to sub-pixel accuracy, which has been discussed in Section 4.2. Blurring may be caused by the optical system (out of focus, aberration etc), relative motion between the imaging system and the original scene, and the point spread function (PSF) of the low resolution sensor (Park et al., 2003). In super resolution reconstruction, the finiteness

of a physical dimension in low resolution sensors is an important factor that affects blur. The low resolution sensor PSF is usually modelled as a spatial averaging operator and the characteristic of the blur is assumed to be known in most cases. The sub-sampling matrix D generates aliased low resolution images from the warped and blurred high resolution image.

Most super resolution image reconstruction methods consist of three stages: registration, interpolation and restoration. At the registration stage, the relative shifts between low resolution frames needs to be estimated with fractional pixel accuracy. In terms of super resolution theory, accurate image registration is a very important factor. Since the shifts between low resolution images are arbitrary, the registered high resolution image will not always match up to a uniformly spaced high resolution grid. Therefore, interpolation is necessary to obtain a uniformly spaced high resolution image from a non-uniformly spaced composite of low resolution images. Finally, image restoration is applied to the un-sampled image to remove blurring and noise (Park et al., 2003). The general process of the super resolution algorithms is illustrated in Figure 4.18.

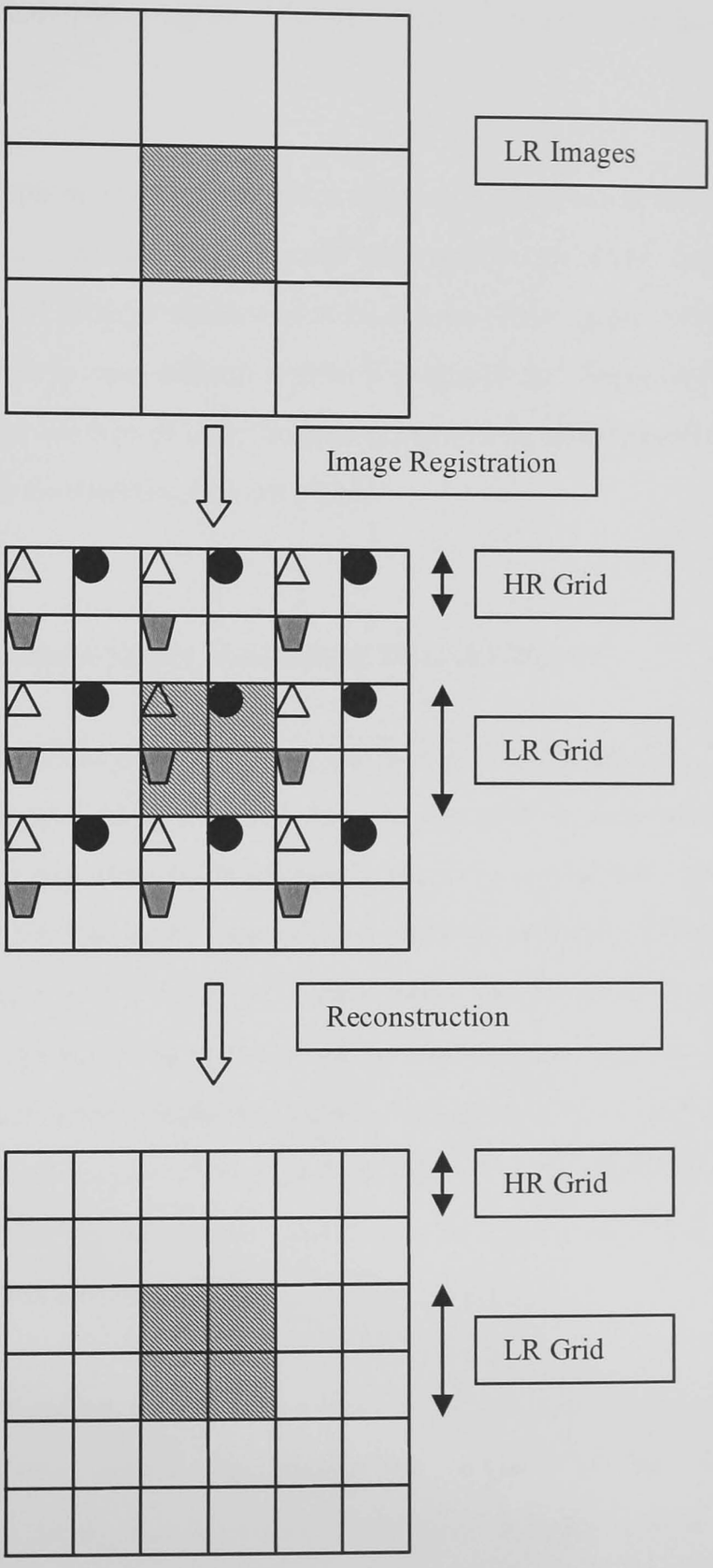


Figure 4.18: An illustration of super resolution reconstruction

4.3.2 Robust Super Resolution Reconstruction on FLIR Images

In most applications, high resolution imagery is more desirable because the higher pixel density provides more useful information. As FLIR images tend to be contaminated by large amounts of noise and are relatively low contrast, detection of small objects is more difficult without post-processing. Super resolution techniques can process this type of noisy image and improve the quality of FLIR imagery so as to improve the detection of small objects.

4.3.2.1 Robust Super Resolution Reconstruction

The super resolution reconstruction idea was first presented by Tsay and Huang (1984). They used a frequency domain approach to demonstrate the ability to reconstruct one improved resolution image from several down sampled noise-free versions of it. Similar frequency domain methods, including a simple version of the above idea applied to noisy and blurred images have also been developed (Park et al., 2003). Frequency based methods have the advantage of being theoretical simple and low computational complexity but are generally difficult to implement in spatially varying degradation models in the frequency domain reconstruction formulation and limited ability fro inclusion of spatial domain *a-prior* knowledge for regularization (Borman and Stevenson, 1998).

The linear spatial domain observation model can accommodate global and non-global motion, optical blur, motion blur, spatially varying PSF (Borman and Stevenson, 1998). The three main tools used for image restoration theory in the spatial domain are the maximum likelihood (ML) estimator, the maximum *a posteriori* (MAP) probability and the set theoretic approach using projection onto convex sets (POCS). Maximum *a posteriori* and projection onto convex sets

algorithms have been the dominant approaches to image super-resolution during the 1990's (Borman, Stevenson, 1998) with recent research focusing mostly on computationally efficient algorithms (Park et al., 2003).

The overlap between the input images can be used to increase the resolution of the FLIR images by super resolution methods. Existing super resolution algorithms solve a very large optimization problem and thus are computationally costly. There is a trend research in super resolution which aims to find effective and fast methods.

Farsiu et al. (2004) have proposed a fast and robust multi-frame super resolution algorithm. They used L_1 norm minimisation and robust regularisation based on a bilateral prior to deal with different data and noise models, with experiments conducted based on visual band LR blurred images. This was developed to reconstruct high resolution images from low resolution images contaminated with non-Gaussian noise. Existing estimation models generally assume a Gaussian noise profile and so can be less effective when the assumptions made on data and noise models do not faithfully relate to the image type. In that case, it is possible that a suboptimal estimation method may produce better and more stable results, since it is less sensitive to modelling and data errors. The L_1 norm minimisation is based on the maximum likelihood method and is robust to non-Gaussian noise. Regularisation is used to de-blur the higher resolution image.

FLIR imagery is not only contaminated by a large amount of non-Gaussian noise, but also has very low contrast. A super resolution reconstruction method should have the essential properties of robustness to outliers and fast implementation. Blurring has not been considered here as there is no explicit optical blurring in the simulated infrared imagery. This simplification does not necessarily limit the performance of the techniques described here because the distortion due to the photon shot noise and non-uniformities in the imager are assumed to be larger than the effect of blurring.

Also, the experiment nature of infrared cameras means that optical blur is minimised and motion blur is small as long as camera motion is relatively small from frame to frame.

A popular family of estimator is the ML-type (M-estimators) (Huber, 1981). Super resolution can be modelled by these estimators:

$$x = \underset{X}{ArgMin}\left[\sum_{k=1}^{N} \rho(y_k, D_k B_k M_k X)\right] \tag{4-11}$$

where ρ is a measure of the distance between the model and measurements.

Considering this model with noise, a least-square approach results in the ML estimate. The least squares formulation is achieved when ρ is the L_1 norm of residual:

$$x = \underset{X}{ArgMin}\left[\sum_{k=1}^{N} \left\| D_k B_k M_k x - y_k \right\|_1^1\right] \tag{4-12}$$

The next step is to reconstruct a high resolution image from a sequence of FLIR images. After aligning by global shift estimation (Fourier method), the FLIR images are up-sampled by a factor of two in the horizontal and the vertical directions in each frame. Subsequently, the normalised cross correlation method can be applied to achieve sub-pixel accuracy. Finally, the super resolution algorithm can be applied to the aligned images. To find the high resolution image x, an L_1 estimator has been used for the minimization problem:

$$x = \underset{X}{ArgMin}\left[\sum_{k=1}^{N} \left\| DM_k x - y_k \right\|_p^p\right], \quad p = 1 \tag{4-13}$$

The gradient of the cost in (4-13) is:

$$d_1 = \sum_{k=1}^{N} M_k^T D^T sign\left(DM_k x - y_k\right) = 0 \tag{4-14}$$

$M_k^T D^T$ copies the values from low resolution grid to the high resolution grid by proper shifting and zero padding, DM_k copies a set of pixels in high resolution grid back on the low resolution grid. Neither of these two operations changes the pixel values. Therefore, the gradient of cost is the aggregation of the effects of all LR frames. When $DM_k x$ is larger than y_k, the sign should be -1, when $DM_k x$ is smaller than y_k, the sign should be +1 and when $DM_k x$ is equal to y_k, the sign should be zero. The zero gradient of cost indicates that there are equal numbers of the minus and positive signs. It means each element of X should be the median value of corresponding elements in the low resolution frames. This method has significant benefits in terms of the computational cost. A flowchart giving an overview of super resolution reconstruction is shown in Figure 4.19.

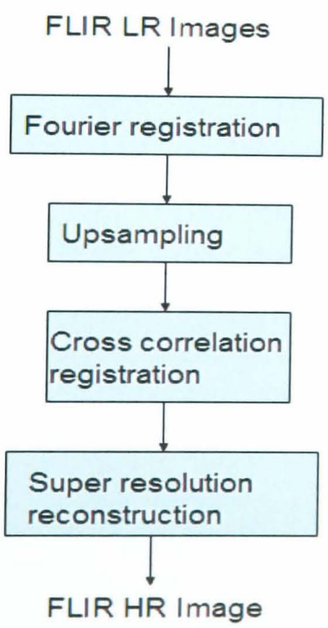
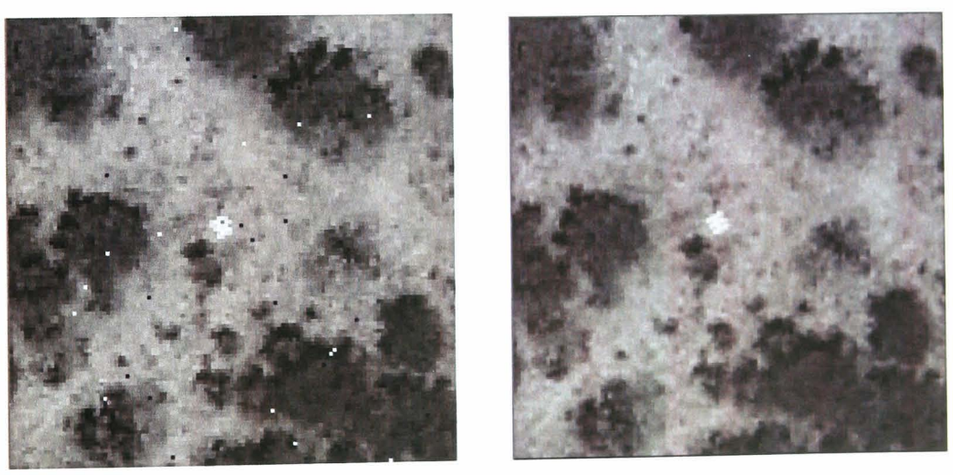


Figure 4.19 An overview of flowchart super resolution reconstruction

4.3.2.2 SR Performance

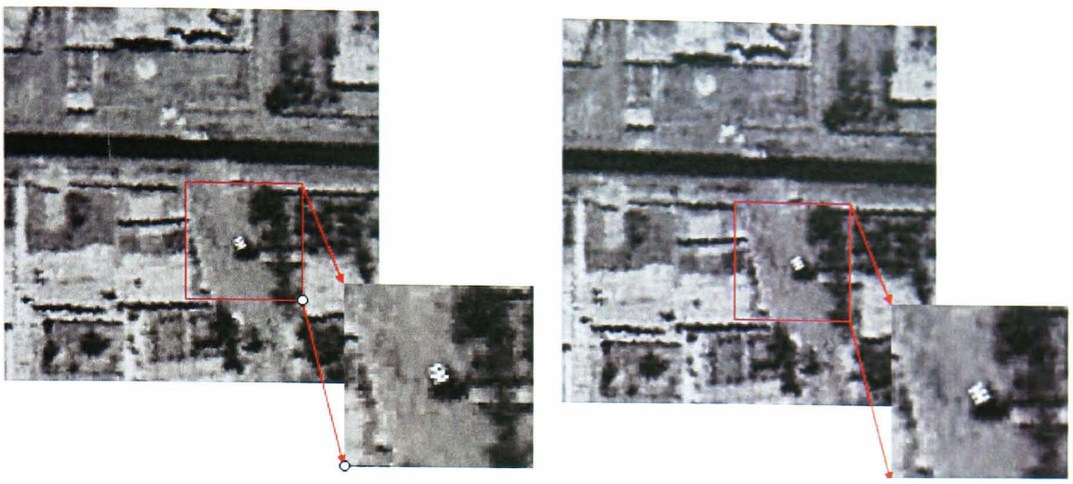
A principal aim of this research is to develop algorithms which work on real FLIR image sequences. However, on several occasions throughout this and later chapters synthetically generated test sequences will be used. This has two benefits over using only real data. First, comparisons can be made between the known ground-truth image and the estimated super-resolution image. Second, factors of the image model which affect the super-resolution estimate can be controlled and investigated in isolation. Such factors include the accuracy of registration, accuracy of the point-spread function, and level of observation noise.

The super resolution reconstruction method has been tested on FLIR images with and without dead/saturated pixels and is shown in Figure 4.20 and Figure 4.21 respectively. The high resolution FLIR image was reconstructed from 25 low resolution FLIR frames.



(a) (b)

Figure 4.20 An example of the super resolution algorithm dealing with noisy images (a) the upsized noisy low resolution image and (b) the super-resolved image by using noisy low resolution images.



(a) (b)

Figure 4.21 Example of (a) One of the low resolution frames and (b) High resolution image formed by 25 low resolution images.

The super resolution algorithm was also tested on FLIR imagery with different background scenes. The super-resolved image for each set was reconstructed from 15 frames of 256 by 256 low resolution images from different imagery sets. The original high resolution and low resolution images were generated by the same image model, except that the image size was different. Figure 4.22 (c), (f), (i), (l) are the super resolved images that the most obvious improvement is around the target area.

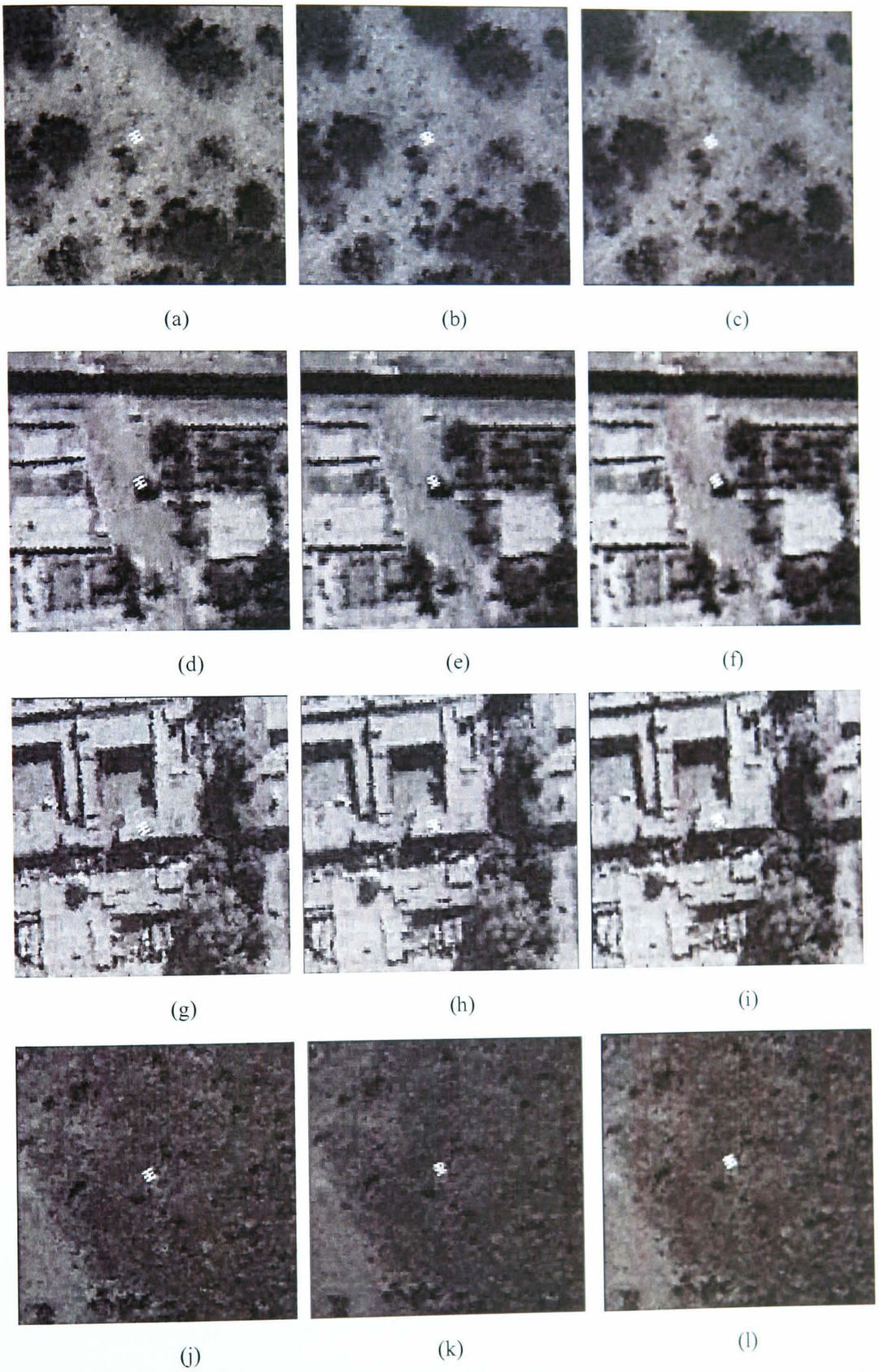


Figure 4.22 Four sets of different imagery with super-resolved results. (a = Pro), (d = Kan), (g = Jal), (j = Tri) are original simulated high resolution images, (b),(e), (h), (k) are the corresponding low resolution images, (c), (f), (i), (l) are corresponding the super-resolved images.

The improvement in the super-resolved images is apparent but still qualitative. Quantitative image improvement is often best illustrated using standard performance metrics, such a cross-correlation or Peak Signal-to-Noise Ratio (PSNR) for the unprocessed low resolution images and the super-resolved image. The performance analysis of the super-resolved images is determined by their similarity to the true high resolution image. There are two different measure approaches employed in this case to compare the results: the Peak Signal-to-Noise Ratio (PSNR) and the Correlation Coefficient (CC) (Chui C. K., 1987 and Wang et al., 2004)

The PSNR method is a commonly used image quality measure because of its low complexity. The value of PSNR is high if the similarity of the images is high. The PSNR is good measure for the reconstructed result of the same image. The Root-Mean-Squared-Error (RMSE) between two images $f(x, y)$ and $g(x, y)$ is:

$$RMSE = \sqrt{\frac{1}{MN}\sum_{n=1}^{M}\sum_{m=1}^{N}[f(x,y)-g(x,y)]^2} \tag{4-15}$$

Then the PSNR is defined as:

$$PSNR = 20 \times \log_{10}\frac{P}{RMSE} \tag{4-16}$$

where *RSME* is the Root-Mean-Square-Error between the original high resolution image and the super-resolved image and P is the maximum pixel value which is 255 for grey level image.

The Correlation Coefficient is the same as the method of cross correlation used for cross correlation image registration and it is used to measure the similarity between two images. If the maximum absolute value of CC is 1, it indicates a perfect match.

The cross correlation values between the first two low resolution images in the sequence with sub-pixel shifts after Fourier registration are shown in Table 4.1. It indicates that the order of correlation values from high to low are: Pro, Kan, Jal and Tri.

Table 4.1 The correlation values of low resolution images with sub-pixel shift

Imagery Sets	CC
Pro	0.9730
Kan	0.9563
Jal	0.9482
Tri	0.9122

To evaluate the effects of the super resolution algorithm, the PSNR and CC are used for measuring the improvement of the ground truth. For low resolution images, the PSNR and CC values are calculated between original simulated high resolution image and resized low resolution (LR) image. To estimate super resolution images, the PSNR and CC values are calculated between original high resolution image and super-resolved (SR) image. The results are shown in Table 4.2. Both PSNR and CC values show positive results, indicating that the image quality is improved after super resolution was applied.

Table 4.2 The PSNR and correlation values for robust super resolution algorithm

Imagery Sets	PSNR (LR)	PSNR (SR)	CC (LR)	CC (SR)
Pro	74.6733	76.0446	0.9547	0.9656
Kan	71.3585	72.4816	0.9391	0.9527
Jal	68.5857	69.8355	0.9020	0.9241
Tri	73.1878	76.8338	0.8634	0.8992

Another way to understand these results is to compute the increase between the similarity value for the low resolution image and the super-resolved image, as shown in Table 4.3. There are some differences between the PSNR and CC methods between different imagery sets.

Table 4.3 The increase values of PSNR and correlation value increase		
Imagery Sets	Difference(Δ)	
	PSNR	CC
Pro	+1.3713	+0.0109
Kan	+1.1231	+0.0136
Jal	+1.2498	+0.0221
Tri	+3.646	+0.0358

For the CC method, it shows that the image Tri is the most improved followed by Jal, Kan and then Pro. The lower the correlation values between the low resolution frames and the original images the larger the improvement can be. Looking at the PSNR results, the Tri image is the most improved image (as the CC method also indicates), followed by the Pro image, and then the Jal and Kan images. This also shows that the simpler background images show greatest improvement in the similarity measure from the application of the super-resolution algorithm.

4.4 Summary

The ability to form a high resolution image from a series of low resolution images is desirable in a variety of military and medical applications. A novel accurate image registration technique, designed specifically for FLIR imagery, suitable for use with the super resolution algorithm has been proposed.

The novelty of this technique is twofold: firstly, this type of image registration technique has never previously adapted for use with FLIR imagery; second, the combination of Fourier transform for coarse image registration followed by normalised cross correlation method for sub-pixel accuracy allows a higher degree of registration accuracy for lower computational cost when compared to normalised cross correlation alone.

This technique has been designed for use with and without the presence of cloud cover. By combining the Fourier registration method with the normalised cross correlation method, the proposed registration method is automatic, robust to outliers, and is accurate to a sub-pixel level.

Next, a robust and effective super resolution algorithm, designed to be effective with FLIR imagery has been proposed. This enhances the resolution of the normally low resolution FLIR images. The L_1 norm method has been proved to be robust to outliers and computationally inexpensive for FLIR imagery.

In next chapter, super resolution algorithms will be applied and enhanced so as to deal with occlusions caused by cloud cover and to reconstruct a broad view of the scene.

Chapter 5 Cloud Effects and Super Resolved Mosaicing on FLIR

5.1 Introduction

The aim of this thesis is to develop a suite of algorithms to automatically detect small objects from cluttered and noisy FLIR imagery. As discussed in chapter 4, when FLIR images are not occluded by clouds, it is relatively easy to reconstruct images by making use of the super resolution algorithm. However, this ideal situation is rare since most of the time FLIR images are covered by certain amount of cloud cover; this adds to the problem of detecting objects that are completely or partially occluded.

Bright clouds can suppress the contrast of the image background and semi-transparent ones can generate multiple ego-motion flow vectors. The ground is the 'background' to the image because the clouds are in the foreground. It is possible to remove most of the effects of cloud cover by processing a sequence of imagery.

In this chapter, an implementation of super resolution has been experimented on FLIR imagery with cloud cover first and then an image mosaicing method has been employed to expand the field of view of the FLIR images. A segmentation method is employed based on optical flow, which is utilised to segment cloud occlusion from the image. An super resolved mosaicing method is presented in this chapter. This method can effectively increase the field of view of a camera by allowing several views of the same scene to be combined into a single high resolution image. It enables object detection be carried out on a large scale of the scene.

5.2 Super Resolution Technique on FLIR Imagery with Cloud

A robust and effective super resolution algorithm has been fully described in chapter 4 and has been tested on synthetic FLIR imagery without the presence of cloud. In this section, the application of the super resolution technique on FLIR imagery with cloud will be considered.

5.2.1 The Effect of Clouds on FLIR Imagery

To achieve super resolution on FLIR imagery with cloud as shown in Figure 5.1, there are three main areas that need be considered: the regions completely occluded by clouds; areas with coverage by partial or semi-transparent cloud; the areas without cloud occlusion. For the regions with full clouds, they can be removed because there is no information that can be obtained about the ground/background. For the areas with coverage by partial or semi-transparent cloud, they can be processed by using contrast enhancement to extract information. For the areas without cloud occlusion, they can be maintained for further super-resolution processing. Therefore, these three types of regions need to be segmented at first.

Figure 5.1 A frame of FLIR imagery with cloud

5.2.2 Cloud Segmentation

Cloud separation of these regions requires an effective segmentation method to segment the different regions of the image. The main goal of image segmentation is to divide an image into parts that have a strong correlation with objects or areas contained in the image (Sonka M. et al., 2008). The segmentation of cloud cover cannot be considered as conventional object segmentation because the cloud does not have clear boundary and its edges become transparent so as to blend in to the background features. For this reason, simple segmentation based on the intensity difference can not be applied.

However, the FLIR imagery is obtained from above cloud which has its own movement that is different from the background movement. By searching for the movement of clouds, its position can be determined. A common means of determining object movement is the optical flow technique. Optical flow is the distribution of apparent velocities of the movement of features (intensity distributions) in an image (Horn and Schunk, 1981). Optical flow takes into account the relative motion of the viewer and objects in the field of view, and segmentation can be achieved because different regions of the same object should have consistent flow vectors. By considering the differences in optical flow vectors different regions can be segmented.

The algorithms used for extracting the optical flow vectors can be classified into gradient-based and feature-based methods. Gradient-based method is adopted here which is based on the same reason for image registration method. Gradient-based techniques rely on the spatial-temporal differential equation describing the motion (Kearney et al., 1987). This can be described mathematically by considering two grey level images of an object, $I(x + dx, y + dy, t + dt) = I(x, y, t)$. The optical flow brightness constraint equation is as:

$$\frac{\partial I}{\partial x}u+\frac{\partial I}{\partial y}v+\frac{\partial I}{\partial t}=0 \tag{5-1}$$

where $u=\frac{dx}{dt}$ and $v=\frac{dy}{dt}$.

For each pixel, a velocity vector indicates the speed of the pixel moving across the image and the direction of its motion. In this research, a block matching algorithm has been employed to obtain the optical flow vector. The image with cloud is divided into overlapping square patches, each approximately 32×32 pixels in size. The reason that each block is overlaid with its neighboring blocks is to make the most of the information in the image. The blocks are registered using the Fourier registration technique described in Chapter 4; the grid of points produces an optical flow field. The optical flow-fields for a cloud covered image are shown in Figure 5.2.

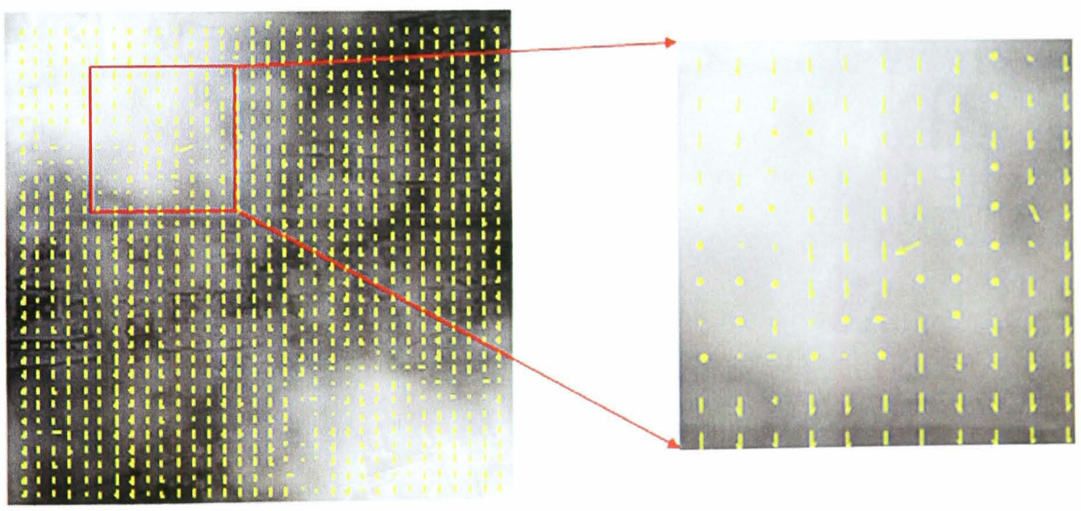


Figure 5.2 Optical flow image

The Fourier registration method registers each block with the corresponding block from the previous frame. Registration provides the shift information in x and y directions for each block in each of the images. This allows the velocity of each pixel to be estimated based on the sequential shifts of the surrounding blocks. The cross-power spectrum of the Fourier transform can be used to indicate the degree of the registration between blocks in adjacent images. For the fully clouded part and the

background, there is no difference in registration – one will provide registration based on the cloud structures (if there are any) and one will register the ground features – however the phase correlation from the semi-transparent part will be suppressed by the cloud.

By drawing the correlation image – a plot of the phase correlation value for each block – the semi-transparent part can be selected by thresholding the phase correlation values. From Figure 5.3, it can be seen that the optical flow vectors change direction around the edge of the cloud. At the interface between the two flow fields, the flow vectors change because of the different flow fields for the clouds and the ground. The lower peak correlation values in this region correspond to the fact that the semi-transparent regions contain multiple flow fields and the correlation for each is suppressed because of the presence of the other. The area, corresponding to the lower correlation value which indicates the semi-transparent cloud area, can be segmented.

As shown in the Figure 5.3, (a) and (b) are a FLIR image without cloud and with cloud cover. Figure 5.3 (c) shows the semi-transparent areas which displayed as red areas. Figure 5.3 (d), (e) and (f) are semi-transparent areas, fully clouded parts and without cloud region respectively.

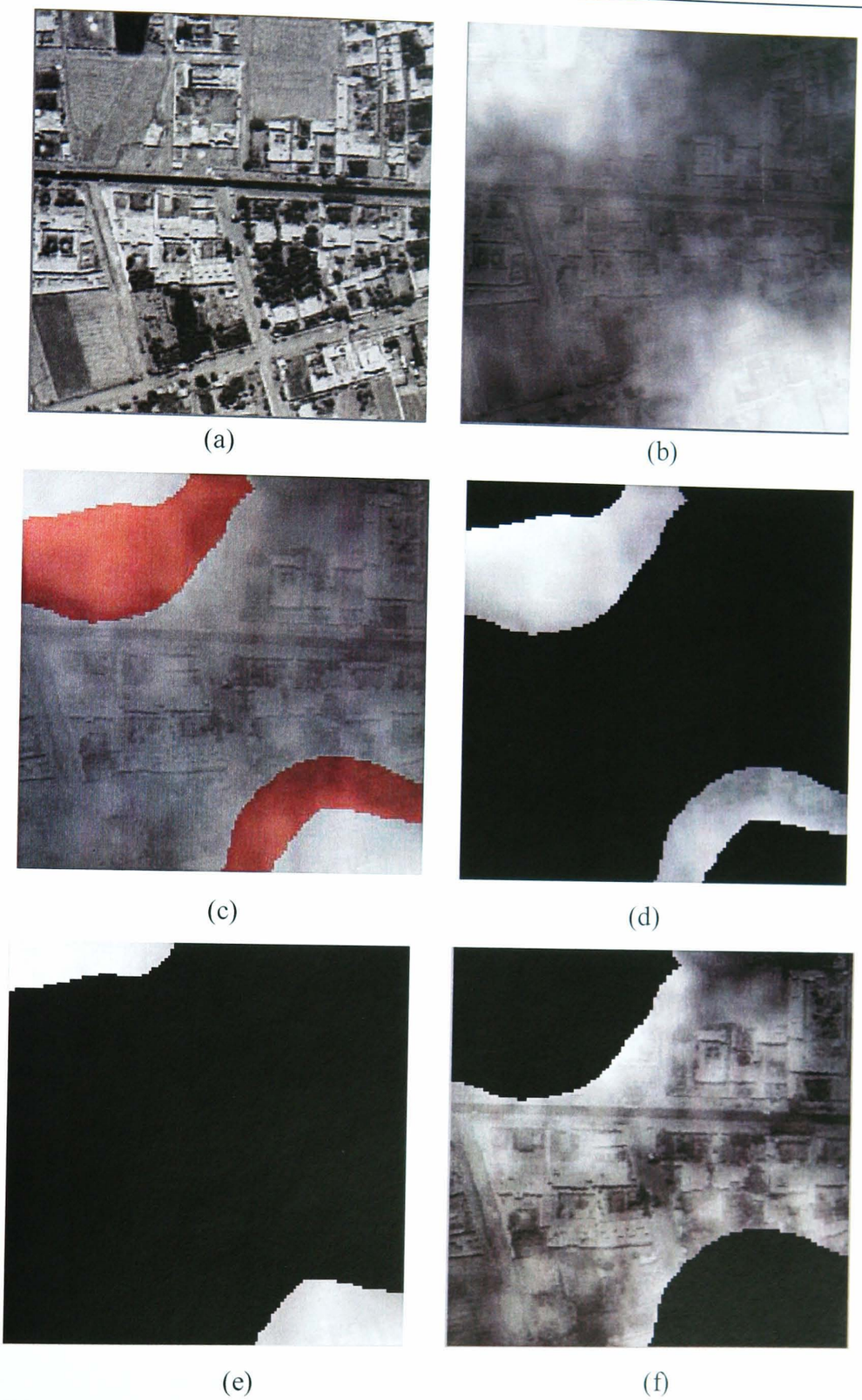


Figure 5.3 Examples of a segmented image for cases with cloud cover: (a) original simulated FLIR image without cloud, (b) original simulated FLIR image with cloud, (c) highlighted parts are the semi-transparent cloud areas, (d) segmented semi-transparent areas, (e) segmented fully clouded parts, and (f) segregated transparent/semi-transparent part of image

5.2.3 Super Resolution Reconstruction on FLIR Imagery with Clouds

The aim of image reconstruction is to recover the desired image from the clouds and enhance the resolution of the image by combining sequential frames. For the semi-transparent cloud areas the background is suppressed, giving features lower relative contrast, but there is still some background information that can be extracted from the image. Therefore, a contrast enhancement method has been employed here to enhance the semi-transparent area and therefore increase the amount of information available.

For the area fully covered by cloud, no background information can be extracted, so it is discarded. For the non-cloud background, this area is kept to be used in enhancing the final image resolution. A block diagram of proposed SR reconstruction algorithm is shown as Figure 5.4.

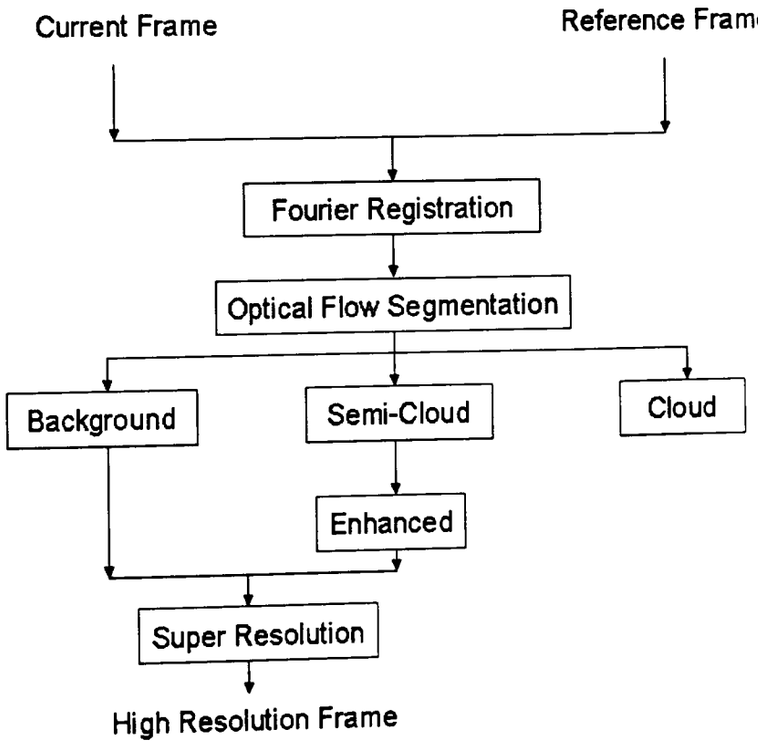


Figure 5.4 Block diagram of proposed SR reconstruction algorithm in presence of clouds

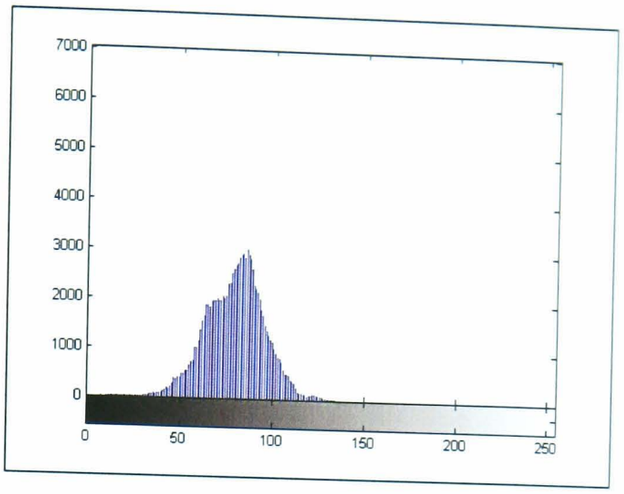
5.2.3.1 Enhancement

A contrast enhancement method has been implemented to improve the visibility of partially occluded regions. Contrast stretching improves the contrast in an image by stretching the actual range of intensity values it contains to span a desired range of values. It differs from the more sophisticated histogram equalization in that it can only apply a linear scaling function to the image pixel values. This results in less harsh enhancement than histogram equalization.

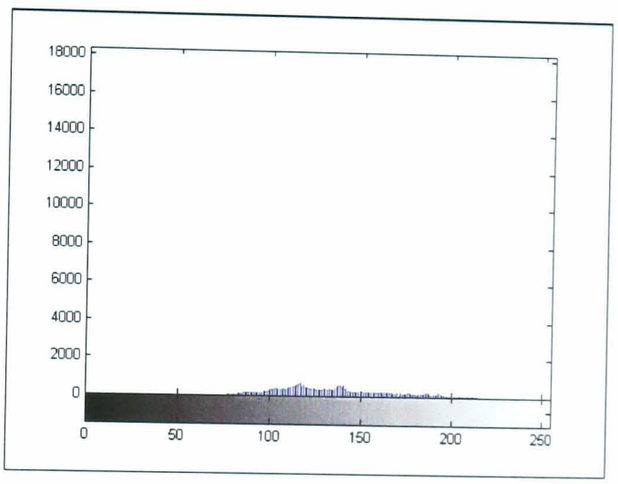
In contrast stretching, it is necessary to specify the upper and lower pixel value limits over which the image is to be normalized. The lower value is the minimum pixel intensity of that image while upper value is the maximum pixel intensity. For a normal 8-bit grey-level image, the lower and upper limits might be 0 and 255. For FLIR images the range is narrower as it is fundamentally low contrast. The aim of the enhancement algorithm is to enhance the semi-transparent cloud area with contrast stretching to reduce the effect of the partial cloud. Given the lower and upper limits - a and b - of a normal contrast image, and the lower and upper limits - c and d - of a low contrast image respectively, each pixel is scaled using following function:

$$P_{out} = (P_{in} - c)\left(\frac{b-a}{d-c}\right) + a \tag{5-2}$$

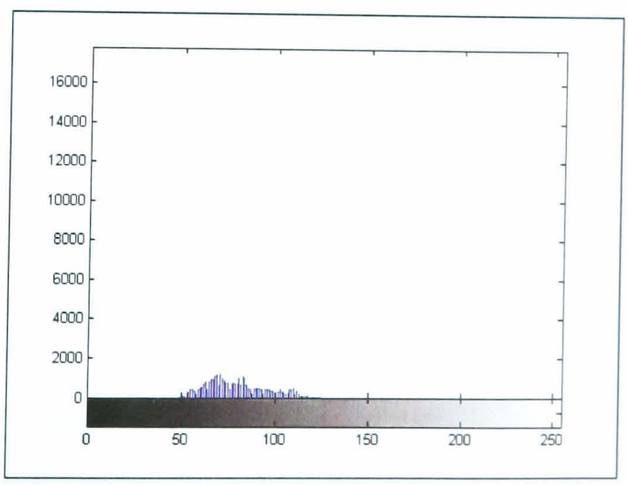
Where *Pout* is the output pixel value and *Pin* is the input pixel value. Values below 0 are set to 0 and values above 255 are set to 255. The results of histogram enhancement is shown as Figure 5.5.



(a)



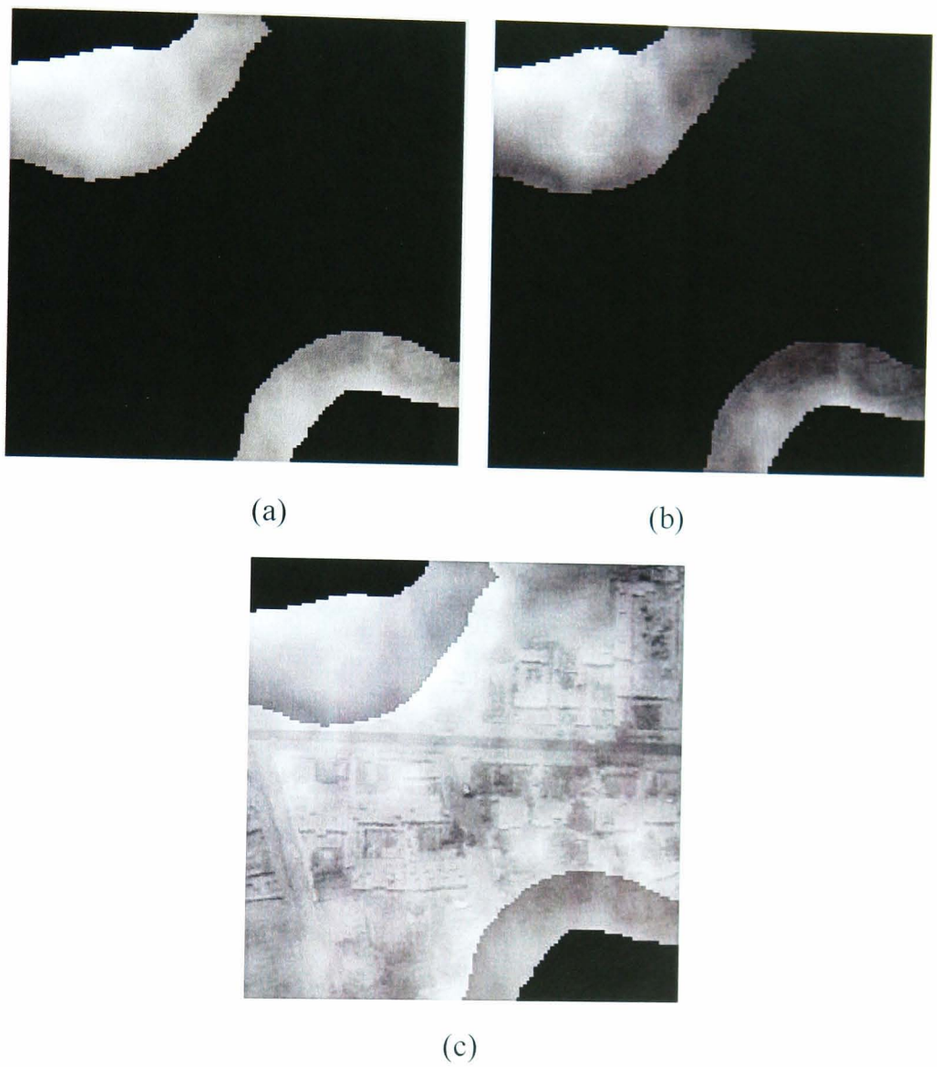
(b)



(c)

Figure 5.5 Histogram of semi-transparent areas (a) histogram for original clear cloud part, (b) histogram for original semi-transparent areas, (c) histogram for enhanced semi-transparent areas.

The semi-transparent areas as shown in Figure 5.6 (a) were enhanced by contrast stretching method. Figure 5.6 (b) is the result after contrast enhancement and Figure 5.6 (c) is the cloud part with enhanced semi-transparent areas.



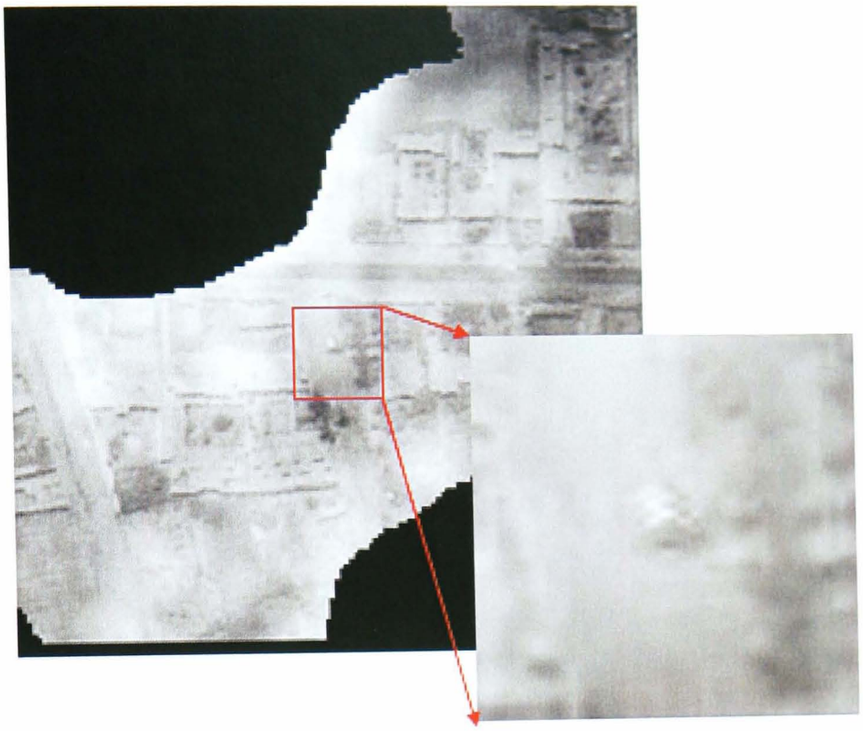
(a) (b)

(c)

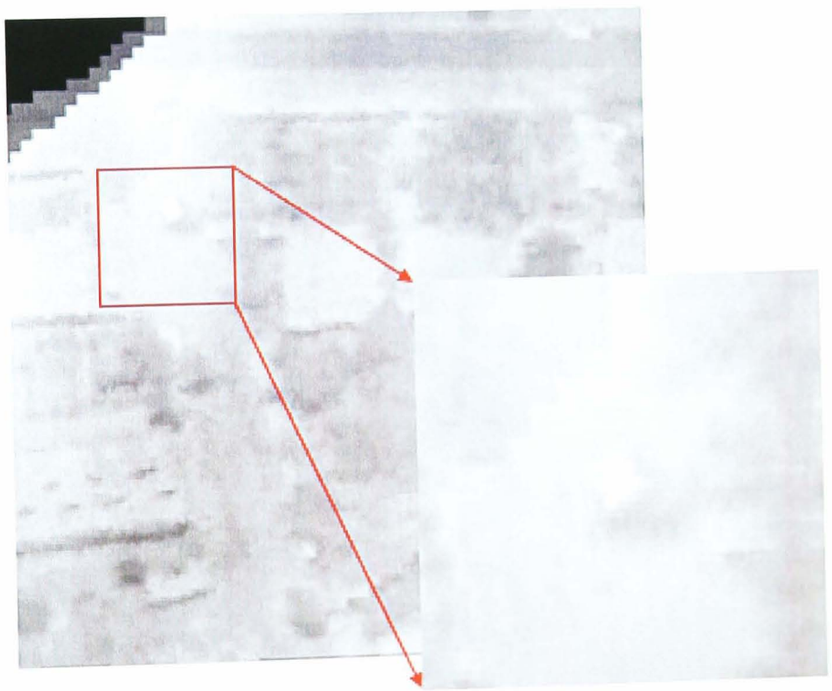
Figure 5.6 Contrast stretching result (a) original semi-transparent areas, (b) enhanced semi-transparent areas, (c) the cloud part with enhanced semi-transparent areas

The super-resolution reconstruction has been applied on the background images only at present. A composite image using the background and the partially cloudy regions can be constructed, but the added information around the target does not currently offer a significant improvement in resolution over such a small area. Figure 5.7 (b) shows the improvement in resolution derived from a sequence of five images, having segmented the cloudy areas using the Fourier registration/phase correlation

techniques described above. The background images were aligned and a sequence of expanded image patches with small targets was generated. Then, the higher resolution image was produced by the efficient and robust super-resolution method described in Chapter 4 and shown in Figure 5.7 (b). It can be seen that the target area has been enhanced by the super-resolution compared with the target in original image. It can provide more information and higher resolution data around the target for further processing, such as target detection and tracking.



(a)



(b)

Figure 5.7 Super resolution reconstruction result: (a) the target area in original image before super resolution, (b) the target area after super resolution.

5.3 Super Resolved Mosaicing Method on FLIR Imagery

To effectively detect objects in a wider area, image mosaicing and super resolution methods can be employed to expand the field of view of the image and enhance the image resolution. In this section, a complete system for super resolved mosaicing of a sequence of FLIR images with some overlap between every two successive images will be presented.

5.3.1 Image Mosaicing Theory

In recent years, image mosaicing synthesis has received substantial attention in both research literatures as well as in the form of commercial applications. A mosaic is a compound image built through properly composing (aligning) a high number of frames and transforming them onto a common reference plane according to some geometric model (Azzari et al., 2005). Mosaicing technique enables to display the information of multiple frames in a single panoramic image as shown in Figure 5.8.

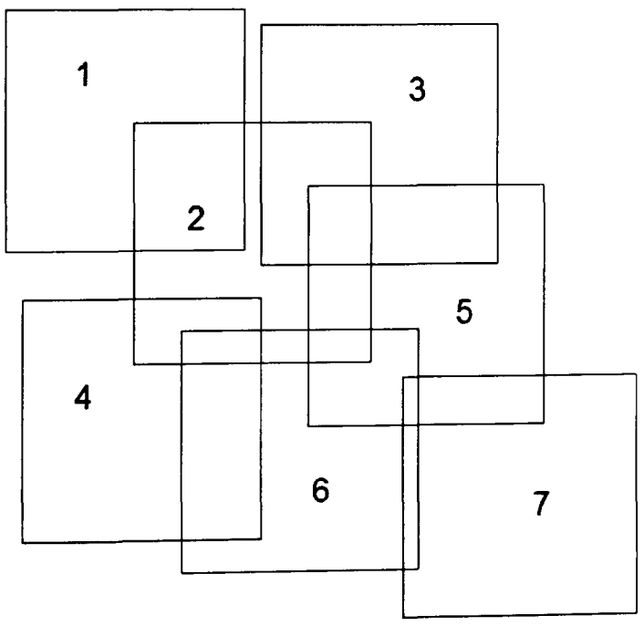


Figure 5.8 An example of creating a mosaicing image from seven overlapping images

Image mosaicing technique enables the creation of larger virtual field of view of camera preserving the original resolution and without introducing undesirable lens deformation. The final mosaicing image is often constructed from a sequence of many images. Similar to super resolution technique, image registration is a primary part of image mosaicing method and accordingly the mosaicing method can be categorized as feature-based methods (Niranjan et al., 2007) and featureless methods by the registration method ((Gouripeddi and Alapati, 2006). As discussed in Chapter 4, global registration method has been proposed to align image accurately.

After registering a sequence of images, the final mosaicing image can be produced by stitching all registered images together. This involves selecting a final compositing surface (flat, cylindrical, spherical, etc.) and view (reference image). It also involves selecting which pixels contribute to the final composite and hot to optimally blend these pixels to minimize the visible seams, blur and ghosting (Szeliski, 2006). Many factors affect the quality of image mosaics, including image alignment, lens distortion, pixel intensity difference among frames, and perspectives of each frame. There is a long history of study on how to create good quality mosaics.

If only a few images are stitched together, a natural approach is to select one of the images as the reference and to then warp all of the other images into the reference coordinate system. The resulting composite is called a flat panorama, since the projection onto the final surface is still a perspective projection, and hence straight lines remain straight. In the case of short range sensors, a flat representation can not be maintained without excessively stretching pixels near the border of the image for large fields of view. In practice, flat panoramas start to look severely distorted once the field of view exceeds 90 ° or so. The usual choice for compositing larger panoramas is to use a cylindrical or spherical. However, since the FLIR imagery is taken from a very long range, the images obtained are maintain on a flat surface. Therefore, a flat surface is generally selected to create a FLIR mosaicing image.

There are two main approaches to stitch image together in the literature, assuming that the images have already been aligned. Optimal seam algorithms search for a curve in the overlap region on which the differences between I_1, I_2 are minimal (Milgram, 1975, Davis, 1998, Efros and Freeman, 2001). Then each image is copied to the corresponding side of the seam. In case the difference between I_1, I_2 on the curve is zero, no seam gradients are produced in the mosaic image I. However, the seam is visible when there is no such curve, for example, when there is a global intensity difference between the images. In addition, optimal seam methods are less appropriate when thin strips are taken from the input images, as in the case of manifold mosaicing (Peleg et al., 2000).

The other approach minimizes seam artefacts by smoothing the transition between images. In Feathering or alpha blending, the mosaic image I is a weighted combination of the input images I_1, I_2. The weighting coefficient (alpha mask) varies spatially as a function of the distance from the seam. The task of blending is to determine the value of a mosaic pixel based on pixels from all warped images. One type of image blending methods creates a mosaic image based on a weighted sum of warped frame pixels. The well known α blending rule is a basic method used for blending frames into a final mosaicing image, according to Equation (5-3) (Azzari et al., 2005),

$$B_{\bar{x},t} = (1-\alpha)B_{\bar{x},t-1} + \alpha I_{\bar{x},t} \tag{5-3}$$

where $I_{\bar{x},t}$ represents a pixel at time t and at position $\bar{x} = (x,y)$. $I_{\bar{x},t}$ is the predicted background value for that pixel and $\alpha \in [0,1]$. Generally, the value of α is selected by experimental evidences. When $\alpha = 0.5$ The simplest weighting function is a flat function that weighs all pixels equally when $\alpha = 0.5$. Image blending by the flat weighting function could lead to visible edges or seams along image boundaries. To

reduce seams, a linear ramp weighting function can be used across the boundary region of two adjacent images. A similar technique (feathering algorithm) was proposed by Shum and Szeliski (1997) to reduce seams. The other blending method is to select the median value of the among the image sequences at each pixel. This type of blending algorithm requires the image alignment is accurate enough; otherwise it is easy to blur the final mosaic image.

5.3.2 Super Resolved Mosaicing Method

With limited optical zoom capability restricted by the system's hardware configuration, SR algorithms provide a promising solution with no additional hardware requirements. Mosaicing techniques can create a panorama image from a sequence of high resolution images. Combining SR algorithm and mosaicing technique together, a super-resolved mosaicing method can be employed to generate a high resolution panorama image.

There are several steps involved in super-resolved mosaicing, such as image registration, super resolution and image mosaicing. Image registration aligns the images onto the same coordinate system and super resolution increase the resolution of the image. After the image has been registered and the resolution has been enhanced, the mosaiced image can be produced. The proposed super-resolved image mosaicing algorithm described here both builds a panoramic image and upgrades the image resolution.

5.3.3 Super Resolved Mosaicing Experiments

It has been shown in Chapter 4 that the globally consistent registration works well on FLIR imagery sequences. For the clear FLIR imagery, the images can be stitching together without considering cloud occlusions. However, when the FLIR imagery is

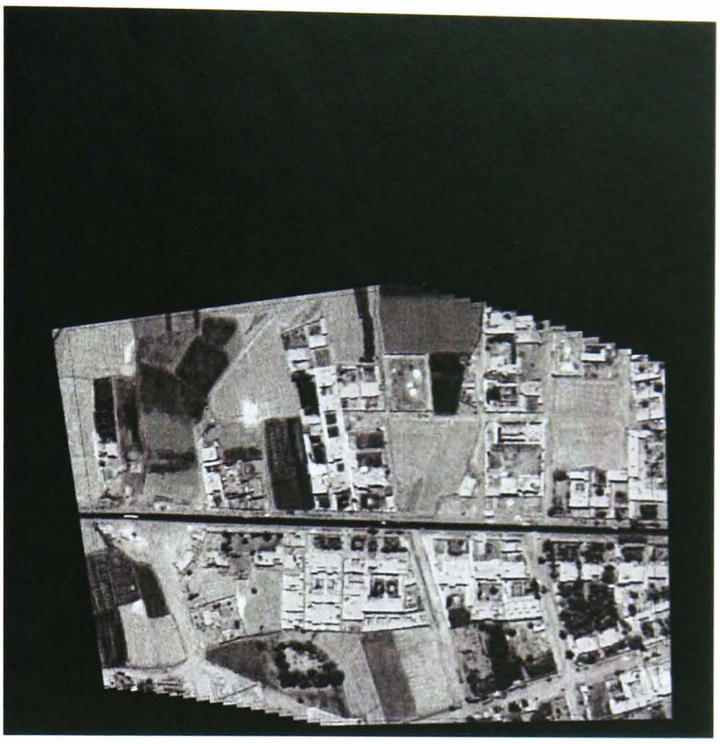
covered by the cloud and the cloud cover is shifted for different frame. They have to be considered differently from the image without cloud cover. The proposed super resolved mosaicing method has been tested on both FLIR images without cloud and with cloud.

5.3.3.1 Super Resolved Mosaicing on Images without Cloud

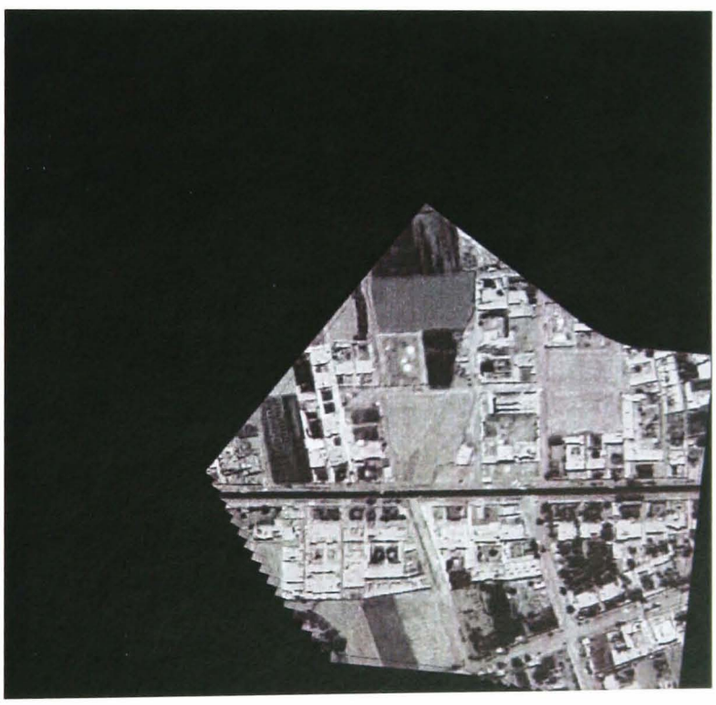
To test mosaicing method on image without cloud cover, there are two sets of FLIR imagery each contained 20 frames have been used. One set is the FLIR imagery with only translational movement and the other is with both translational and rotational movement.

The results of super resolved mosaicing images from two sets of FLIR imagery are displayed as Figure 5.9. Figure 5.9 (a) is high resolution mosaiced image from 20 low resolution FLIR images without cloud and only with translational movement. Figure 5.9 (b) is the high resolution mosaiced image from 20 low resolution FLIR images without cloud and with translational and rotational movement.

Both results have shown that the super resolved mosaicing method has successfully increased the image resolution and enlarge the field of view of the original low resolution images.



(a)



(b)

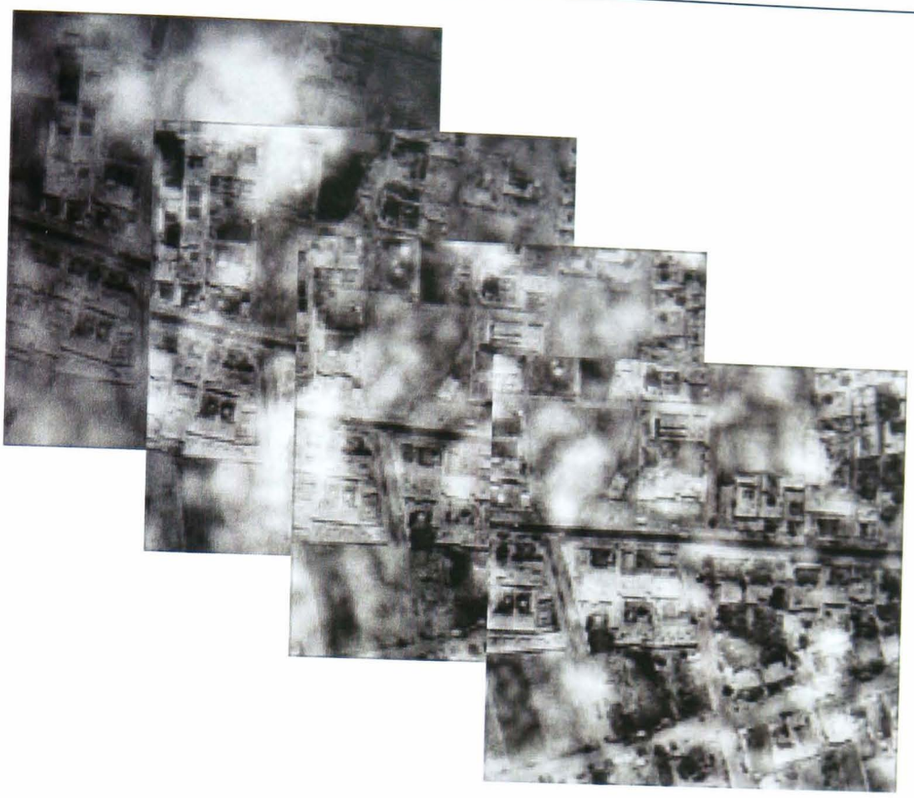
Figure 5.9 (a) Super resolved mosaicing result of 20 low resolution FLIR images without cloud (without rotation), (b) Super resolved mosaicing result of 20 LR FLIR images without cloud (with rotate angle 2 degree for each frames)

5.3.3.2 Super Resolved Mosaicing on Images with Cloud

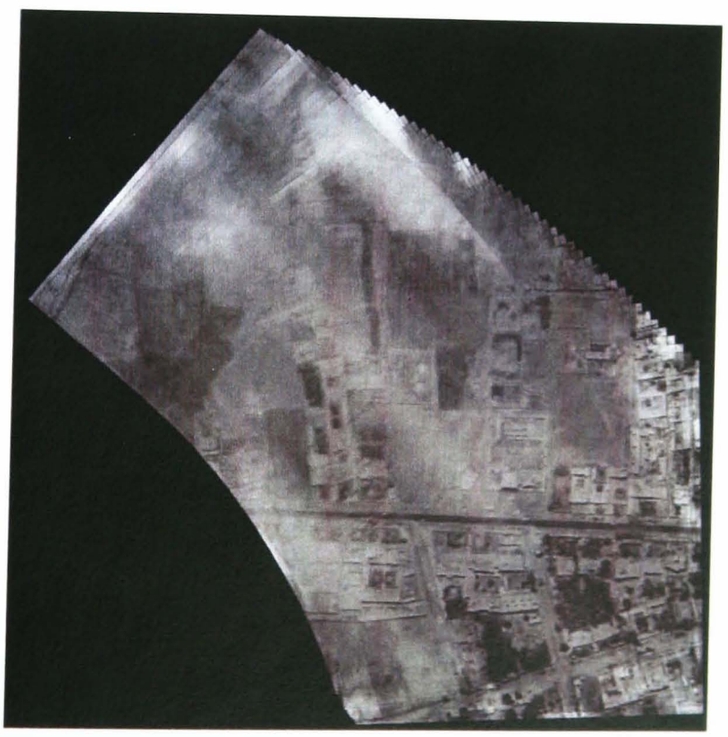
As shown in Figure 5.10, the clouds are scattered over the images and moved in each frame in the sequence. In order to test super resolved mosaicing method on FLIR imagery with cloud, there are two ways can be applied. As proposed in Chapter 4, super resolved method has been applied only on the segmented clear part of FLIR imagery. It is applicable on a short sequence of image when all segmented clear frame have large enough overlapped area to register. However, when dealing with a large amount of frames, it is difficult to register all the segmented frames. Therefore, it has been considered to apply the super resolved mosaicing method directly on the translated and rotated low resolution FLIR imagery with cloud.

The results of super resolved mosaicing images from two sets of FLIR imagery are displayed as Figure 5.9. Figure 5.9 (a) is high resolution mosaiced image from 20 low resolution FLIR images without cloud and only with translational movement. Figure 5.9 (b) is the high resolution mosaiced image from 20 low resolution FLIR images without cloud and with translational and rotational movement.

As show in Figure 5.10, the result indicates that the proposed super resolved mosaicing method can successfully generate the super-resolved mosaiced image in the presence of the cloud cover. This method is robust to the cloud cover and can effectively remove most of the cloud cover from the image.



(a)



(b)

Figure 5.10 (a)An example of the frames in the low resolution image sequences with cloud, (b) the mosaicing results from a sequence of FLIR images with cloud (Rotate angle 2 deg)

5.4 Summary

This chapter has described the development of a combined super resolution and mosaicing method which builds upon the techniques discussed in Chapter 4. It has been designed specifically for use on noisy and low contrast FLIR imagery.

The proposed super resolved mosaicing method has the novelty of adapting a combination of techniques more commonly used on visible band imagery for use in FLIR. Super resolution allows the assessment of a specific area by increasing image resolution, whereas mosaicing is used to develop situational awareness at a less specific level. By forming a high resolution image using super resolution in the first instance, image mosaicing is more effective in allowing a more accurate situational awareness.

Images with and without cloud cover have been considered. Images with could cover are more complex to process due to the requirement to consider the movement and occlusion caused by the cloud. Two methods to overcome these challenges have been investigated. A segmentation method based on optical flow has been proposed, with the intention of segmenting the cloud by using its movement, whilst leaving the transparent and semi-transparent areas to allow further processing. The second technique involved applying the super resolved mosaicing method directly to images with cloud cover and was found to have a smoothing effect over the frames.

Overall, it has been shown that super resolved mosaicing can be applied to images with and without cloud cover, to enhance resolution in the first instance, which then improves situational awareness following image mosaicing.

Chapter 6 Review of Object Detection

6.1 Introduction

Object detection has played a crucial role in computer vision in the past decades. The goal of object detection is to determine the regions that are highly likely to contain objects of interest. Given an object class of interest T (the target) and an image I, object detection is the process of detecting and locating the occurrences of T in I (Zhang et al., 2005). The most common approach for object detection is to use the information in a single frame. On the other hand, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections (Alper et al., 2006). This temporal information is usually in the form of frame differencing, which highlights changing regions in consecutive frames.

For FLIR imagery, the target is normally buried in the noise or cluttered background. Many approaches have been proposed for object detection in cluttered images. These algorithms can generally be categorized into two classes: global appearance-based approaches and component-based approaches. Global appearance-based approaches take an object as a unit and perform classification on features generated from the entire object (Zhang et al., 2005). Component-based methods treat an object as a collection of parts. These methods first extract some object components, and then detect objects by using geometric information (Zhang et al., 2005). As the objects to be detected are small in size, possibly smaller than the noise that is presented in the imagery, there are several global-based methods can be used for this problem.

6.2 Reviews of Detection Algorithms

The term object detection here refers to the detection of small objects in large images. This could include both object classification and object localisation. Object classification refers to the task of discriminating between images of different kinds of

objects, where each image contains only one of the objects of interest. Object localisation refers to the task of identifying the positions of all objects of interest in large image. The object detection problem is crucial step for some image processing tasks such as automatic target recognition and automatic object recognition. In this part, a general literature review has been carried on image thresholding, image differencing and background modelling which might be used in the future project.

6.2.1 Image Thresholding

Object detection can be considered a segmentation process which aims to segment the stationary or moving object from the image. Segmentation is the process of separating objects from background. Therefore, image thresholding can be considered as a basic detection algorithm.

In application where specify grey values of regions are not important, a picture can be segmented into "object" and "background" by simple choose a threshold in brightness. The region above the threshold is object and all below the threshold as background. There are several ways to choose thresholds, ranging from the trivially simple to the very sophisticated. As the sophistication of the technique increases, performance improves but at the cost of increased computational complexity. A single threshold is almost never appropriate for an entire scene. It is always the local contrast between object and background that contains the relevant information. When a single threshold cannot provide sufficient performance, the local thresholds have to be chosen. The most common approach is called block thresholding, in which the picture is partitioned into rectangular blocks and different thresholds are used in each block. Typical block size can be 32×32 or 64×64 for a 512×512 images. The block is first analyzed and a threshold is chosen, then that block of image can be used for further analysis.

6.2.2 Image Differencing

Image differencing has been proposed as another efficient ways for object detection. It is an image processing technique that can be used to determine changes between images. The difference between two images is calculated by finding the difference between each pixel in each image, and generating an image based on the result. For this technique to work, the two images must first be aligned so that corresponding points coincide, and their photometric values must be made compatible, either by careful calibration, or by post-processing. The complexity of the pre-processing needed before differencing varies with the type of image.

Hsieh et al. (2006) proposed a method for detection of small object with low contrast. As noise which means dead and saturated pixels presented in small objects images usually drastically affects the detection performance because it is difficult to distinguish noise from small object, they proposed a novel noise removal technique (Hsieh et al., 2006). According to their method, the difference image of the two images f_t and f_{t+1} is defined as

$$d_t(x,y) = |f_t(x,y) - f_{t+1}(x,y)| \tag{6-1}$$

An optimal threshold value can be determined to discriminate the background clutter from foreground objects by assuming that noise pixels possess relatively low difference values, and the noise is assumed to be randomly distributed in the background and statistically independent. So it can be defined that

$$f(x,y) = s(x,y) + n(x,y) \tag{6-2}$$

where $f(x,y)$ is the observed value in a 2-D image, $s(x,y)$ is the real signal, and

$n(x,y)$ is the noise signal. Mostly, noise can be assumed to be zero-mean Gaussian distribution with variance σ^2. As the noise pixels in two contiguous images f_t and f_{t+1} are both captured from a video camera, there are two properties of the noise can be obtained: the number of noise pixels must be larger than that of targets and the noise should be much more scattered than the objects. The noise model for images f_t and f_{t+1} is assumed to be a Gaussian distribution, characterized by two independent random variables X_1 and X_2. Finally, Hsieh et al. got the conclusion that the number of non-zero neighbours surrounding a specific pixel (x,y) is attained via the following equation:

$$S(x,y) = \left| \{(i,j) | (i,j) \in N_{x,y}, d(i,j) > 0 \} \right| \tag{6-3}$$

where $N_{x,y}$ is the neighbourhood of pixel (x,y).

Next, the function that counts the pixel number having the same number of neighbours is defined as

$$M(s,t) = \left| \{(x,y) | (x,y) \in D, S(x,y) = s, d(x,y) = t \} \right| \tag{6-4}$$

and s is the number of non-zero neighbours of the pixel, D is the difference image. So a histogram can be obtained according to the following formula:

$$H(t) = \sum_{s=0}^{k} M(s,t), \quad t = 1, 2, 3, \ldots. \tag{6-5}$$

where k is the maximal number of noise neighbours. The threshold value T in determining noise and non-noise pixels is set up to be the argument corresponding to

the maximum absolute value in a differentiating histogram $H(t)$. The following formula summarizes the procedure:

$$T = \arg\left(\max_{t>0} \left| \frac{\partial H(t)}{\partial t} \right| \right) \tag{6-6}$$

This image differencing algorithm can get good results on the difference image with noise. But in some cases if the contrast of the objects is too low, the objects' contours may not be successfully extracted.

6.2.3 Background Modelling

Object detection can be achieved by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame (Alper et al., 2006). The process sometimes is referred to as background subtraction. Background subtraction is widely used as a basis of motion detection from image sequences. In traditional background detection algorithms, the pixel values of the current frame are compared with the corresponding pixel values of the fixed reference background image (Ji et al., 2006). With this approach it is possible to detect new objects in the scene even if they suddenly stop moving and detect objects removed from the scene. A simple background model usually assumes that the background pixels are static over time. The foreground objects can then be obtained by subtracting the current frame from the background image. It is simple, but extremely sensitive to changes in dynamic scenes derived from lighting and extraneous events etc. Therefore, it is highly dependent on a good background model to reduce the influence of these changes, as part of environmental modelling (Hu et al., 2004).

Many works have been proposed in the literature as a solution to an efficient and

reliable background subtraction. As far as back 1970s Wren et al. (1997) modelled the colour of each pixel, $I(x,y)$, with a single 3D (Y, U, and V colour space) Gaussian,

$$I(x,y) \sim N(\mu(x,y),\xi(x,y)) \tag{6-7}$$

The model parameters, the mean $\mu(x,y)$ and the covariance $\xi(x,y)$, were learned from colour observations in consecutive frames. Once the background is derived for every pixel (x, y) in the input frame, the likelihood of its colour coming from $N(\mu(x,y),\xi(x,y))$ is computed, and the pixels that deviate from the background model are labelled as the foreground pixels. This work by Wren et al is one of the earliest works related to background modelling. Mittal and Huttenlocher (2000) have created a model of the wide field of view that can be used to distinguish between the static background and the moving objects, or foreground. They modelled the scene using a mixture model for each of the pixel locations in the mosaic. The probability of a pixel belonging to a particular Gaussian is proportional to the weight ascribed to that Gaussian. The probability of a pixel having an intensity value x which belongs to a particular Gaussian j can be simplified as

$$P(X = x) = \sum_{j=1}^{K} \omega_j P(X = x | J = j) \tag{6-8}$$

This scheme is useful in modeling multi-modal backgrounds and scene changes as well as modelling transient moving objects. Cho and Kim (2005) proposed a statistical background mosaic model have some advantages especially in the case of surveillance systems in which the camera moves with a regular pattern. To reduce accumulated errors, they estimated the global motion (GM) between the current image and the estimated background mosaic which contain no objects instead of the

previous image. After GM compensation, they assume a generally spatio-temporal statistics of the background model

$$X(x,y,t) \sim N(m_t(x,y,t-1)\xi_t(x,y,t-1)) \tag{6-9}$$

$$[X(x,y,t) - m(x,y,t-1)] \sim N(0,\xi_s t) \tag{6-10}$$

where $X(x,y,t)$ is the extracted feature from the current image which uses the intensity and gradient at the same time. The disadvantage of this algorithm is that misdetection errors may be accumulated and propagated through the whole sequence as the background mosaic is updated after each one detection cycle. As the normal background subtraction is extremely sensitive to illuminations changes, El Maadi and Maldague (2007) proposed a novel dynamic background subtraction technique which takes into account variations in speed of scene illumination. They combined edge detection with a binary foreground image computed with a dynamic threshold which led an improved segmentation quality with higher robustness. Their approach is based on an effective background subtraction method: the current frame is subtracted from the scene's estimated background image and the result is thresholded providing foreground objects:

$$|Frame_i(x,y) - Bg(x,y)| > Th \tag{6-11}$$

where $Frame_i(x,y)$ is the current frame at time i, $Bg_i(x,y)$ is the estimated background images at the same time i, and Th is the threshold value which is applied to generate a binary image of the objects of interest. They use a temporal low-pass filter to update the estimated background image as:

$$Bg_{i+1} = Bg_i + \alpha \times (Frame_i - Bg_i) \quad (6\text{-}12)$$

where α is a learning rate set by the user. With a constant value ofα, the gradual illumination changes can be taken into account but it fails when subject to suddenly quick and strong variations in illumination. To overcome these problems, they use an adaptive gain-threshold based upon the illumination history which can control and adjust the gain-threshold combination dynamically according to the speed of the illumination change.

In summary, there are many new algorithms related to object detection have been proposed in the last six years, image differencing, background modelling are among the most popular methods. Although most state-of-the-art objects detection methods for fixed cameras use background subtraction methods to detect regions of interest, they can provide a good starting point for a moving camera scene. There are gradually more and more new algorithms dealing with moving cameras and illumination changes are developed which can be applied in the future project.

Chapter 7 Conclusions and Future Work

7.1 Thesis Summary

The work presented in this thesis has been concerned with enhancing FLIR image quality so as to improve object identification and detection accuracy. This has applications in several different fields, particularly in the military environment.

FLIR images are low contrast and suffer from high noise levels, which presents challenges for small object detection and identification. These issues stem from the inherent limitations of the FLIR system and so image processing is the main means of improving image quality.

Previous work has considered the super resolution method to be an effective means of increasing the resolution of low resolution images. However, a robust and fast method is required for use on FLIR imagery. In Chapter 4, a sub-pixel image registration method was proposed. This sub-pixel registration method is a combination of the Fourier registration and cross correlation methods enabling sub-pixel accuracy whilst being computationally efficient. Accurate sub-pixel image registration then allowed super resolution to be applied more effectively to reconstruct a high resolution image from the sequence of FLIR images.

The image mosaicing method is widely used to expand the field of view from a sequence of images allowing more objects to be detected in a single process. The FLIR super-resolved mosaicing method developed in this thesis can not only expand the field of view but also increase the resolution of images. This method has been designed to be especially suitable for FLIR imagery and is robust to outliers and so can improve the quality of the image and therefore improve detection performance.

7.2 Directions for Future Work

In this thesis, investigations have been carried out on four research areas: image registration, super resolution, mosaicing and object detection algorithms.

Compared to conventional image detection algorithms, the key point in this thesis was to enhance image quality so as to improve the performance of subsequent detection algorithms. A system, summarised in Figure 7.1, has been designed in previous chapters to increase the resolution and expand the field of view of FLIR images. In future work, different detection algorithms could be applied so as to evaluate the ability of this research to improve automatic detection functions.

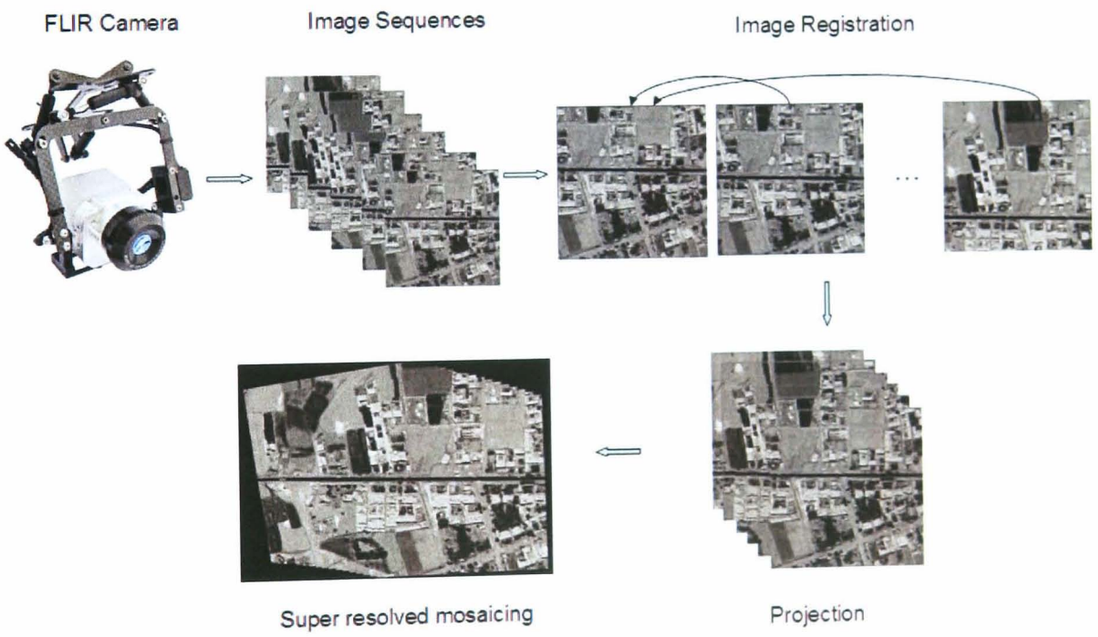


Figure 7.1 A system for improving the performance of detection algorithms

This thesis focused on the detection of static objects. This system could be expanded for moving object detection. An issue that arises in scenes with moving objects is that since the moving object translates with respect to the background, a mosaic consistent with every image in the sequence is not possible. Standard compositing techniques blend all available information and produce a blurred image in moving

regions. Although a theoretically correct mosaic is not possible when objects in the scene are moving, a typical user is only interested in a plausible reconstruction.

Further investigation could be carried out on the errors in motion estimation so as to improve accuracy and reliability. Estimation of image transformation parameters can be biased by moving objects because moving objects in the image indicate a transformation different than the transformation due to the camera movement.

REFERENCES

ABIDI B. R., ARAGAM N. R., YAO Y., ABIDI M. A., 2008, Survey and Analysis of Multimodal Sensor Planning and Integration for Wide Area Surveillance, ACM Comput. Surv., Vol. 41, No. 1., pp. 1-36.

Abraham B. N., Arecchi T. F., Lugiato A. L., 1988, Instabilities and chaos in quantum optics II, North Atlantic Treaty Organization. Physics, Vol. 177.

Abramowitz, M. and Stegun, I. A., 1972, Interpolation. §25.2 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 878-882.

Anon., 2002, AN03: Guide to image quality and pixel correction methods, Rad-icon Imaging Corp, pp1-4.
(**http://www.rad-icon.com/pdf/Radicon_AN03.pdf#search=%22AN03%22**)

Adiv G., 1985, Determining three-dimensional motion and structure from optical flow generated by severl moving objects, IEEE Trans Pattern Anal Mach Intell 7(4):384-400.

Alam M.S., Bognar J.G., Hardie R.C., and Yasuda B.J., 2000, Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames, IEEE Trans. Instrum. Meas., vol. 49, pp. 915-923.

Alper Y., Omar J., Mubarak S., 2006, Object tracking: A survey, ACM Computing Surveys, vol. 38(4).

Anuta P.F., 1969, Digital registration of multispectral video imagery, Soc. Photo-

Optical Instrum. Engs. Journal, vol. 7, pp. 168-175.

Arfken, G., 1985, Development of the Fourier Integral, Fourier Transforms--Inversion Theorem and Fourier Transform of Derivatives, §15.2-15.4 in Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press. pp. 794-810.

Argyriou, V., Vlachos, T., 2004, Using gradient correlation for sub-pixel motion estimation of video sequences, IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings. vol.3, pp. iii - 329-332.

Azzari, P., Di Stefano, L., Bevilacqua, A., 2005, An effective real-time mosaicing algorithm apt to detect motion through background subtraction using a PTZ camera, IEEE Conference on Advanced Video and Signal Based Surveillance, 2005, pp. 511 – 516.

Barnea D. I., Silverman H. F., 1972, A class of algorithms for fast digital image registration, IEEE transactions on computers, vol. c-21, NO. 2, pp. 179-186.

Bates J.R. and Gregory W. D., 2000, Voice & data communications handbook, McGraw-Hill Osborne Media.; 3 Sub edition, pp. 170-171

Bevilacqua A. and Azzari P., 2007, A Fast and Reliable Image Mosaicing Technique with Application to Wide Area Motion Detection, Image Analysis and Recognition, Lecture Notes in Computer Science, Volume 4633/2007, pp. 501-512.

Bilodeau, G.A., Torabi, A., Morin, F., 2011, Visible and infrared image registration using trajectories and composite foreground images, Image & Vision Computing, Vol. 29, Issue 1, pp.41-50.

Borman S., Stevenson R., 1998, Spatial Resolution Enhancement of Low-Resolution Image Sequences A Comprehensive Review with Directions for Future Research,

Brown L. G., 1992, A survey of image registration techniques, ACM computing surveys (CSUR), vol. 24, Issue4, pp. 325-376.

Bulanon, D.M., Burks, T.F., Alchanatis, V., 2009, Image fusion of visible and thermal images for fruit detection, Biosystems Engineering, Vol. 103, Issue 1, pp.12-22.

Burns N. R., Butavicius M., Sunde J., Hanton K., 2010, Super-resolution of Infrared Images: Does it Improve Operator Object Detection Performance?, Journal of Computing & Information Technology

Calafiore G. C., 2000, Outliers robustness in multivariate orthogonal regression, IEEE Trans. Syst., Man. Cybern., vol. 30, no. 6, pp. 674–679.

Capel D. and Zissserman A., 1998, Automated mosaicing with super-resolution zoom. In proceedings of the conference on computer vision and pattern recognition, Santa Barbara, pp. 885-891.

Castleman K.R., 1996, Digital Image Processing, Prentice-Hall, Englewood Cliffs, NJ, pp 161-163, 187-188,200.

Chalermwat P., 1999, High performance automatic image registration for remote sensing, Thesis, George Mason University.

Cho J. H., Kim S. D., 2005, Object detection using multi-resolution mosaic in image sequences, Signal Processing: Image Communication, v 20, n 3, pp. 233-253.

Chaudhuri S., 2002, Super-Resolution Imaging, Springer US, Electronic resource.

Heidelberg: Springer-Verlag GmbH, pp.73-105.

Chen H., Varshney P. K., and Arora M. K., 2003, Performance of mutual information similarity measure for registration of multitemporal remote sensing images, IEEE Trans. Geosci. Remote Sens., vol. 41, no. 11, pp. 2445–2454.

Chui C. K., 1987, Wavelets: A Mathematical Tool for Signal Analysis, Society for Industrial & Applied Mathematics, U.S., pp. 179-180.

Cole-Rhodes, A.A., Johnson, K.L., LeMoigne and J., Zavorin, I., 2003, Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient, Image Processing, IEEE Transactions on Image Processing, Vol. 12, No. 12, pp. 1495 – 1511.

Dani P. and Chaudhuri S., 1995, Automated assembling of images: Image montage preparation, Pattern Recognition 28, pp. 431-445.

De Castro E. and Morandi C., 1987, Registration of translated and rotated images using finite Fourier transforms, IEEE Transactions on Pattern Analysis and Machine Intelligence archive, Vol. 9, Issue 5, pp. 700-703.

Debella-Gilolow M., and Kääba A., 2011, Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation, Remote Sensing of Environment, Vol. 115, Issue 1, pp. 130-142

Driggers R. G., Richard H. V., 2000, Analysis of Sampled Imaging Systems, SPIE Press, pp148-149.

Driggers R., Sept 2003, Encyclopedia of Optical Engineering: Vol 2, Taylor & Francis Inc, pp1466.

Dvornichenko V.N., Bounds on (Deterministic) correlation function with application to registration, IEEE trans. Pattern Anal. Machine Intell., vol.5, no.2, pp. 206-213, 1983.

El Maadi A., Maldague X., 2007, Outdoor infrared video surveillance: A novel dynamic technique for the subtraction of a changing background of IR images. Infrared Physics and Technology, vol. 49, n 3, SPEC. ISS., pp. 261-265.

Elad M. and Feuer A., 1997, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images,IEEE Trans. Image Processing, vol. 6 , no. 12, pp. 1646-1658.

Farsiu S., Robinson D., Elad M. , and Milanfar P., 2004, Advances and challenges in super resolution, the International Journal of Imaging Systems and Technology.

Fisher R. B., Dawson-Howe K., Fitzgibbon A., Robertson C., Trucco E., Dictionary of computer vision and image processing, John Wiley & Sona, Ltd, pp58.

Fonseca L.M.G., Manjunath B.S., 1996, Registration techniques for multisensory remotely sensed imagery, Photogrammetric Engineering and remote sensing 62, pp. 1049-1056.

Foroosh H. and Hoge W. S., 2003, Motion information in the phase domain, in Video registration, M. Shah and Kumar R. (EDs), Kluwer Academic Publishers, Boston. MA 36-71.

Galileo S., 2007, Long wave 2D MCT Detectors: An overview of LW detector technologies, pp1-4.

Gao, Z., Gu, B., Lin, J., 2008, Monomodal image registration using mutual information based methods, Image and Vision Computing; vol. 26 (2), pp.164-173

Gonzalez R. C., Woods R. E., 2006, Digital Image Processing (3rd Edition), Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Goshtasby, A., Stockman, G. C., Page, C. V., 1986, A region-based approach to digital image registration with subpixel accuracy, IEEE Transactions on Geoscience and Remote Sensing, vol. GE-24, 1986, pp. 390-399.

Goshtasby A., 1995, Egde detection by curve fitting, Image and Vision Computing, Vol. 13(3), pp. 169-177.

Goshtasby A., 2005, 2-D and 3-D image registration for medical, remote sensing, and industrial applications, Hoboken, NJ : J. Wiley & Sons.

Gouripeddi P. R., Alapati M., 2006, A Novel Still Image Mosaicing System Using Featureless Registration, Binary Check Stitching and Minimal Blending, Image processing, computer vision and pattern recognition, IPCV'06, CISST INTERNATIONAL CONFERENCE, pp. 1:223-229.

Gross H., Zügge H., Peschka M., and Blechinger F., 2007, Handbook of Optical Systems: Aberration Theory and Correction of Optical Systems vol. 3.

Halliday D. and Rensnick R., Physics, 1978, 3rd edn (new york: Wiley), pp.1035.

Haralick R. and Shapiro L., 1993, Computer and Robot Vision, Addison-Wesley, Reading, MA.

Harris, C.J., Stephens, M., 1988, A combined corner and edge detector. In: Proc. 4th

Alvey Vision Conferences, pp. 147–151.

Henini M., Razeghi M., Manijeh M. R., 2002, Handbook of Infra-red Detection Technologies, Elsevier Science & Technology, pp38-39.

Hsieh F.Y., Han C.C., Wu N.S., Chuang T.C., Fan K.C., 2006, A novel approach to the detection of small objects with low contrast, Signal Processing, v 86(1), pp71-83.

Holst G. C. and Lomheim T. S., 2007, CMOS/CCD Sensors and Camera Systems, Winter Park, FL. and Bellingham, WA: JCD Publishing and SPIE Optical Engineering Press, pp 191.

Horn B. and Schunk B., 1981, Determing Optical Flow, Artificial Intelligence 17(1-3), pp.185-203.

Hough P. V. C., 1962 Method and means for recognizing complex patterns, U.S. Patent 3, 069,654.

Howe D. J., 1993, The Infrared and Electro-Optical Systems Handbook, Environmental Research Institute of Michigan and The Society of Photo-Optical Instrumentation Engineers, Volume 4, page 92.

Hrka´c T., Kalafati´c Z., and KrapacJ, 2007. Infrared-Visual Image Registration Based on Corners and Hausdorff Distance, Proceeding SCIA'07 Proceedings of the 15th Scandinavian conference on Image analysis.

Hu W., Tan T., Wang L., Maybank S., 2004, A survey on visual surveillance of object motion and behaviors, IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, v 34(3), pp. 334-352.

Huang T. S. and Tsay R. Y., 1984, Multiple frame image restoration and registration, in Advances in Computer Vision and Image Processing, vol. 1, T. S. Huang, Ed. Greenwich, CT: JAI, pp. 317–339.

Huber P. J., 1981, Robust Statistics. New York: Wiley.

Ibarra-Castanedo C., 2005, Quantitative subsurface defect evaluation by pulsed phase thermography: depth retrieval with the phase, Ph. D. thesis, Laval University, pp. 128.
Available at : **http://www.theses.ulaval.ca/2005/23016/23016.pdf**
[Accessed 2 December 2009].

Irani M., Anadan P., 1998, A unified approach to moving object detection in 2D and 3D scenes. IEEE Trans Pattern Anal Mach Intell 20(6):577-580.

Itakura Y., Tsutsumi S., Takagi T., 1974, Statistical Properties of the Background Noise for the Atmospheric Windows in the Intermediate Infrared Region, Infrared Physics, Vol. 14, pp17–29.

Jahne B., 1997, Digital image processing: concepts, algorithms and scientific application, 4th edition, Springer.

Ji X.P., Wei Z., Feng Y., 2006, Effective vehicle detection technique for traffic surveillance systems, Journal of Visual Communication and Image Representation, vol. 17(3), June, Real-Time Imaging, pp. 647-658.

Kearney J.K., Thompson W.B., Boley D.L., 1987, Optical flow estimation: an error analysis of gradient based methods with local optimization, IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.9, No. 2, pp. 229-244.

Keller, Y., Averbuch, A., Israeli, M., 2005, Pseudopolar-based estimation of large translations, rotations, and scalings in images, IEEE Transactions on Image Processing, Vol. 14 (1), pp. 12-22.

Kenney, J. F. and Keeping, E. S, 1962, The Standard Deviation and Calculation of the Standard Deviation, §6.5-6.6 in Mathematics of Statistics, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand, pp. 77-80.

Kim B. K., Kim S. J., Choi S. J., 2008, Fourier Based Image Registration for Sub-Pixel Using Pyramid Edge Detection and Line Fitting, Intelligent Networks and Intelligent Systems, International Workshop, pp. 535-538.

Kirill I. K., 1969, Radiation in the atmosphere, International geophysics series, Academic Press, Vol. 12, pp. 189.

Klein S., Pluim J. P. W., Staring M., Viergever M. A., 2009, Adaptive Stochastic Gradient Descent Optimisation for Image Registration, International Journal of Computer Vision, Vol. 81, Issue 3, pp. 227-239.

Kopp C., n.d., Thermal imaging sensors, Defence Today NCW 101 Series Articles part 6, pp. 53.
Available at: **www.ausairpower.net/NCW-101-6.pdf**
[Accessed 5 September 2009].

Krapels K., Driggers R. G., Murrill S., Schuler J., Thielke M., and Young S. S., 2004, Superresolution performance for undersampled imagers,in Defense and Security Symposium (Formerly AeroSense), Proc. SPIE 5407, pp.139–149.

Le Moigne J., Campbell W.J., Cromp R.F., 2002, An automated parallel image registration technique based o the correlation of wavelet features, Geoscience and

Remote Sensing, IEEE Transactions, Vol. 40, No. 8, pp. 1849 -1864.

Leachtenauer J. C., Driggers R. G., 2001, Surveillance and reconnaissance imaging systems: modeling and performance Prediction, pp. 163-164.

Lee J.H., Kim Y.S., Lee D., Kang D., and Ra J.B., 2010, Robust CCD and IR image registration using gradient-based statistical information, IEEE signal processing letters, Vol. 17, No. 4, pp. 347-350.

Leese J. A., Novak S. G., Clark B. B., 1971, An automatic technique for obtaining cloud motion from geosynchronous satellite data using cross correlation. A. Applied Meteorology, Vol. 10, pp. 110-132.

Lehmann T. M., Gonner C., Spitzer K., 1999, Survey: interpolation methods in medical image processing, IEEE Transactions on Medical Imaging 18, pp.1049-1075.

Lewis J. P., 1995, Fast Template Matching, Vision Interface 95 Canadian Image Processing and Pattern Recognition Society, pp. 120-123.

Li Q., Sata I. and Murakami Y., 2007, Efficient stochastic gradient search for automatic image registration, Int J simul model 6(2), pp.114-123.

Li S.S., Liu H. C. and Tidron M. Z., 1998, Long Wavelength Infrared Detectors and Arrays, Physics and Applications, Boston, Mass, pp. 1-11.

Liu R., 2009, Eigentargets versus kernel eigentargets: detection of infrared point targets using linear and nonlinear subspace algorithms, Springer Science + Business Media, Journal of Infrared, Millimeter and Terahertz Waves, Vol. 30, No.3, pp 278-293.

Lloyd J. M., 1975, Thermal Imaging Systems, Springer, 1st edition, pp.166-183.

Lucas B. D., Kanade T., 1981, An Iterative Image Registration Technique with an Application to Stereo Vision, proceedings of imaging understanding workshop, pp. 121-130.

Maes F., Collignon A., Vandermeulen D., Marchal G., and Suetens P., 1997. Multimodality image registration by maximization of mutual information, IEEE Trans. Med. Imag., vol. 16.

Maintz J. B. A. and Viergever A. M., 1998, A Survey of Medical Image Registration, Medical Image Analysis, Vol. 2, pp. 1-37.

Manduchi R. and Mian G. A., 1993, Accuracy Analysis for Correlation-Based Image Registration Algorithms, Circuits and Systems, , ISCAS '93, 1993 IEEE International Symposium, vol.1, pp. 834 – 837.

Mather P. M., 2004, Computer processing of remotely sensed images: an introduction (3rd edition), John Wiley & Sons, pp9-10.

Maurer T., Wilson D.L., Smith S.R., Deaver D.M., Flug E.A., and Nguyen O.T., 2009, Search and detection comparing midwave and longwave infrared, Optical Engineering,vol. 48(11), SPIE, 116401.

McAndrew A., 2004, Introduction to digital image processing with matlab, Course Technology, pp 143-183.

McCann H., Scott D.M., 2005, Process imaging for automatic control, CRC Press, 1 edition, pp.47.

McGuire M., 1998, An image registration technique for recovering rotation, scale

and translation parameters, NEC Tech. Report.

Mikolajczyk K., Schmid C., 2004, Scale & Affine Invariant Interest Point Detectors, International Journal of Computer Vision 60(1), pp. 63–86

Miller J. L., 1994, Principles of Infrared Technology: a practical guide to the state of the art, New York : Van Nostrand Reinhold.

Mittal A., Huttenlocher D., 2000, Scene modeling for wide area surveillance and image synthesis, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, v 2, pp. 160-167.

Modersitzki J., 2004, Numerical methods for image registration, Oxford: Oxford University Press.

Mooney J.M., Shepherd F.D., 1996, Characterizing IR FPA nonuniformity and IR camera spatial noise, Infrared Physics and Technology 37, pp 595-606.

Niranjan, S. Gupta, G. Mukerjee, A. Gupta, S., 2007, Efficient Registration of Aerial Image Sequences Without Camera Priors, Computer Vision – ACCV 2007, Volume 4844/2007, Springer Berlin / Heidelberg, pages 394-403.

Pan W., Qin K., and Chen Y., 2009, An Adaptable-Multilayer Fractional Fourier Transform Approach for Image Registration, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 3, pp. 400-413.

Pan Z. and Atungulu G. G., 2011, Basic Laws of Radiative Heat Transfer, Infrared Heating For Food and Agriculture Processing, Available at: **http://www.ceramicx.com/en/infrared-heat**.

Park C. S., Park K. M., Kang G. M., 2003, Super-resolution image reconstruction: a technical overview, Signal Processing Magazine, IEEE, Vol. 20(3). pp. 21 – 36.

Parker J.A., Kenyon R.V., Troxel D.E., 1983, Comparision of interpolating methods for image resampling, IEEE Trans. Med. Imag., vol. 2, no.1, pp. 31-39.

Peleg S., 1981, Elimination of seams from photomosaics, Proc. Conf. Pattern Recognition and Image Processing, pp. 426–429.

Peleg S., Rousso B., Rav-Acha A., and Zomet A., 2000, Mosaicing on adaptive manifolds. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(10):1144–1154.

Pratt W. K., 1974, Correlation Techniques of Image Registration, IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-10 Issue 3, pp. 353-358.

Pratt W.K., 1991, Digital Image Processing, Wiley, New York, NY.

Rajan D. and Chaudhuri S., 2001, Simultaneous estimation of super-resolved intensity and depth maps from low resolution defocused observations of a scene, in Proc. IEEE int. Conf. Computer Vision, Vancouver, Canada, pp. 113-118.

Ralph J. F., Bernhardt M., 2002, Smart Imaging in the Infrared, *Contemp. Phys.*, vol. 43, pp. 259-272.

Ralph, J. F., Smith, M. I., Heather, J. P., 2005, Motion-based detection, identification. and tracking for missile warning system applications, PROCEEDINGS of SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, 5809:53-64.

Reddy S. B. and Chatterji B. N., 1996, An FFT-Based Technique for Translation.

Rotation, and Scale-Invariant Image Registration, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 5, NO. 8, pp. 1266-1271.

Rosenfeld A., Kak A. C., 1982, Digital picture processing, Academic Press, New York, 2nd ed.

Rybicki, G. B., Lightman, A. P., 1979, Radiative Processes in Astrophysics. New York: John Wiley & Sons.

Roshni V.S., Revathy K., 2008, Using mutual information and cross correlation as metrics for registration of images, Journal of theoretical and applied information technology, Vol. 4, Issue 6, pp. 474-481.

Sarvaiya, J.N.; Patnaik, S.; Bombaywala, S., 2009, Image Registration by Template Matching Using Normalized Cross-Correlation, International Conference on Advances in Computing, Control, & Telecommunication Technologies, pp. 819-822.

Sawhney H. S. and Kumar R., 1999, True Multi-Image Alignment and its Application to Mosaicing and Lens Distortion Correction, IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 21(3), 235-243.

Schoenberg I.J., 1969, Cardinal interpolation and spline functions, Jounal of Approx. Theory, vol. 2, pp. 167-206.

Shum H. and Szeliski R., 1997, Creating full video panaromic image mosaics and environmental maps, Proc. SIGGRAPH, pp. 251–258.

Shum H.-Y. and Szeliski R., 2000, Construction of panoramic mosaics with global and local alignment, International Journal of Computer Vision, Vol. 36(2), pp101-130.

Siegwart R., Nourbakhsh I., 2004, Introduction to autonomous mobile robots. MIT Press, pp. 118-122.

Singh I., 2009, Design of infrared optical system, ICOP 2009-International Conference on Optics and Photonics, India.

Smith B. C., 1995, Fundamentals of Fourier Transform Infrared Spectroscopy, Taylor & Francis Ltd, pp44-46.

Snyder W. E., Qi H., Machine vision, 2004, Cambridge : Cambridge University Press, pp. 298-309.

Sonka M., Hlavac V. and Boyle R., 2008, Image processing analysis, and machine vision 3rd edition, Cole Publishing Company.

Sterel A. and Aggarwal J.K., 1999, Detection moving objects in airborne forward looking infrared sequences, IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications, pp.3.

Stewart V.C., 1999, Robust parameter estimation in computer vision, Society for Industrial and Applied Mathematics, Vol. 4 , Issue 3, pp. 513 – 537.

Stone, H.S., Orchard, M.T., Ee-Chien Chang, Martucci, S.A., 2001, A fast direct Fourier-based algorithm for subpixel registration of images, IEEE Transactions on Geoscience and Remote Sensing, Vol. 39(10), pp. 2235 – 2243.

Strehl A., Aggarwal J.K., 2000, MODEEP: a motion-based object detection and pose estimation method for airborne FLIR sequences, Machine Vision and Applications. Vol. 11, No. 6, pp. 267-276.

Strehl A., Aggarwal J.K., 2000, MODEEP: a motion-based object detection and pose estimation method for airborne FLIR sequences, machine vision and applications, vol.11, No.6, pp 267-276.

Szeliski, R., 2006, Image alignment and stitching: a tutorial, Handbook of mathematical models in computer vision, Springer, pp273-292.

Szeliski R. and Coughlan J., 1997, Spline-Based Image Registration, International Journal of Computer Vision, 22(3), pp 199-218.

Taylor J. R., 1997, Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, University Science Books, U.S., Ed. 2nd, pp.102-103.

Thévenaz P. and Unser M., 2000, Optimization of mutual information for multi-resolution image registration, IEEE Trans. Image Processing, vol. 9, pp.2083–2099.

Tian Q. and Buhns M.N., Algorithms for subpixel registration, Computer Vision. Graphic and Image Processing, 35, pp. 220-233, 1986.

Tidrow M.Z., Dyer W.R., 2001, Infrared sensors for ballistic missile defense, Infrared Physics & Technology 42, p333-336.

Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., Stathaki, T., 2010, Robust FFT-Based Scale-Invariant Image IEEE Transactions on Registration with Image Gradients, Pattern Analysis and Machine Intelligence, Vol. 32 (10), pp. 1889-1906

Unser M., Aldroubi A., and Eden M., 1995, Enlargement or reduction of digital images with minimum loss of information, IEEE Trans. Image Processing, vol. 4, no. 3, pp. 247-258.

Vandewalle P., Süsstrunk S., and Vetterli M., 2006, A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution, EURASIP Journal on Applied Signal Processing, Vol. 2006.

Vincent, E., Laganiere, R., 2002, An empirical study of some feature matching strategies. Proc. Conf. Vision Interface, Calgary, Canada, pp. 139–145.

Viola P., Wells W.M., 1997, Alignment by maximization of mutual information, International Journal of Computer Vision, vol. 24, pp.137–154.

Wang W.H. and Chen Y.C., 1997, Image registration by control points pairing using the invariant properties of line segments. Pattern Recognition Letters 18, pp. 269-281.

Wang Z., Bovik A. , Sheikh H., and Simoncelli E., 2004, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Processing, pp. 1-14.

Weisstein E. W., 2010, Standard Deviation, From MathWorld, A Wolfram Web Resource. Available at: **http://mathworld.wolfram.com/StandardDeviation.html** [Accessed 5 Jan 2010]

Wells III W. M., Viola P., Atsumi H., Nakajima S., and Kikinis R., 1996, Multi-modal volume registration by maximization of mutual information, Med. Imag. Anal., vol. 1, pp. 35–51.

Winkelman F., Patras I., 2004, Online globally consistent mosaicing using an efficient representation, 2004 IEEE international conference on systems, man and cybernetics, pp. 3116-3121.

Wren C. R., Azarbayejani A., Darrell T., Pentland A. P., 1997, Pfinder: real-time tracking of the human body, IEEE Transactions on Pattern Analysis and Machine

Intelligence, v 19(7), pp. 780-785.

Yang C., 2001, Registration of SAR and FLIR images for ATR applications, Proc. SPIE of Signal Processing, Sensor Fusion, and Target Recognition X, Ivan Kadar, Vol. 4380, pp. 127-134.

Young S. S. and Driggers G. R., 2006, Superresolution image reconstruction from a sequence of aliased imagery, Applied optics, Optical society of America, Vol.45, No. 21, pp. 5073-5085. (very good one)

Zappe H., 2010, Fundamentals of Micro-Optics, University of Freiburg, Germany.

Zhang H.M., Gao W., Chen X., Zhao D., 2005, Learning informative features for spatial histogram-based object detection, Proceedings of the International Joint Conference on Neural Networks, v 3, Proceedings of the International Joint Conference on Neural Networks, IJCNN 2005, pp. 1806-1811.

Zhang Z., Blum R. S., 2001, A hybrid image registration technique for a digital camera image fusion application, Information Fusion, Vol. 2, No. 2, pp. 135-149.

Zibaeifard M., Rahmati M., 2006, An Improved Multi-Stage Method for Medical Image Registration Based on Mutual Information, Proceedings of the international conference on computer and communication engineering, pp. 27-32.

Zitova B., Flusser J., 2003, Image registration methods: a survey, Image and vision computing 21, pp. 977-1000.