THE UNIVERSITY
OF ADELAIDE
AUSTRALIA

SUB CRUCE LUMEN

A Study on Image Change Detection Methods for Multiple Images of the Same Scene Acquired by a Mobile Camera

Guntur Tanjung

School of Mechanical Engineering
The University of Adelaide
South Australia 5005
Australia

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy in
Mechanical Engineering
on the 12th October 2009

CHAPTER 1

INTRODUCTION

The research presented in this dissertation concentrates on a study of automated image change detection methods for multiple indoor and outdoor images of the same scene captured by a mobile camera from slightly different positions, angles and at different times. The following sections describe a motivation which drives this research, problems, objectives and scopes of this research, and structure of this dissertation.

1.1 Background and Motivation

Large protected areas, both civilian and military, such as airfields and the perimeters of defence bases are commonly marked by physical barriers known as wire fences (Fig. 1.1). Outdoor perimeter security is regularly patrolled by human guards whose task is to check whether there are any intruders and/or any anomalous objects within the protected areas and whether there are any breaches in the integrity of fence wires (e.g., holes). In increasing the security level of perimeter security, high-cost and high-technology electric fence systems are normally deployed in outdoor perimeter security (Rich, 2007). These electrified fence systems are often composed by numerous Close-Circuit Televisions (CCTVs) coupled to motion detection systems or infrared (IR) beams and taut wire systems (Senstar-Stellar and Magal Security Systems, 2007). Another option in protecting a large protected area is by installing only a low-cost non-electric fence system. In addition to human guards, it would be useful to have an automated system (e.g., a patrol robot equipped with various sensors and an automated sensor data processing system) whose niche would be to assist human guards. Thus, a robust sensor data processing system (e.g., an image processing system for visual and/or IR images) will be required for such robotic application.

NOTE:
This figure is included on page 2
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 1.1 An airport marked by a chain-link mesh fence. The wire fence is used in preventing intruders to get into the airport

A robot, guided by a Global Position System (GPS), could be tasked in patrolling large protected areas. How a patrol robot might traverse a large protected area is described in the brief scenario that follows: The robot commences its patrol from point A. At point A, it halts for a specified period (e.g., 2 seconds) to acquire an image of the area, referred to as a reference image. After capturing the reference image, the robot continues its patrol to the next image acquisition point (e.g., point B) and so on, until the entire area has been traversed. The robot will theoretically have returned to point A but will not be exactly on point because of navigation errors caused by robot internal sensors like GPS, compass and accelerometer tolerances and the terrain condition. This scenario illustrates that an automated image processing system is required for such robotic application and it must be capable of dealing with multiple outdoor images of the same area containing fence wires acquired by a mobile camera from slightly various viewing positions, angles and at diverse times.

In outdoor perimeter security, electric fence systems can detect human motion, while ignoring nuisance events triggered by rain, wind or animals. Perimeter sensors employed in the electrified fence systems can be divided into two broad categories; (1) line sensors and (2) volumetric or area-coverage sensors (Rich, 2007). Line sensors involve physical contact with the sensor in order to trigger an alarm or beam-break sensors. Line sensors consist of taut wire, vibration, acoustic, IR and contact barrier sensors. Volumetric sensors secure a defined area in which a target's motion, presence or absence may be detected. Area-coverage sensors include radio frequency, microwave, electrostatic field, video motion detection (VMD), seismic and ground radar sensors. Although electric fence systems can be employed to protect large areas, they are very expensive in costs including a visibility study prior to install them, installation, calibration and maintenance.

A scanning camera approach can be applied to trace fence wires (Haering, *et al.*, 2008). The approach consists of specified waypoints visited at specified times, speeds and zoom-levels. The approach uses Pan-Zoom-Tilt (PZT) cameras. Although successive frames may contain visible structures of wires, they may not unique enough to provide positional information of significant changes, such as breaches in the integrity of fence wires and/or attached objects in front of fence wires, because of motion estimates from PZT cameras and controlling software are not standardized, tiny sizes of fence wires and non-uniform illumination that occurs along fence wires during the sunny and overcast days.

A patrol robot guided by a GPS can be deployed in outdoor perimeter security. The patrol robot must be able to navigate autonomously and to detect automatically any significant changes that happen in perimeter security. However, an automated image analysis system that can detect presence and absence of objects behind fence wires, breaches in the integrity of fence wires and objects left behind by intruders in multiple outdoor images of the same scene, in which the scene is marked by a wire fence, taken by a mobile camera from somewhat dissimilar viewing positions, angles and at varying times is still uninvestigated yet. Referring to the literature and the author's knowledge, this research is an initial study in automated change detection methods for multiple outdoor images of the same scene containing fence wires captured by a mobile camera from slightly different viewing positions, angles and at different times.

1.2 Problems, Objectives, and Scopes

Detecting significant changes (e.g., absence and presence of objects, objects motion relative to the background and changes of objects in shape) while rejecting unimportant changes caused mainly by camera motion, sensor noise, illumination variation and non-uniform attenuation (Radke *et al.*, 2005) in multiple images of the same scene acquired by a mobile camera is a complex task. Complexity further increases when change detection is performed in an outdoor scene (i.e., a lot of background clutter and a significant day-time variation in illumination) that contains fence wires (i.e., wires of the fence can separate a sizeable object into many small objects. Chain mesh fences are often composed of tiny and thin metal substance. As a result, specular reflections apparently appear on fence wires during the sunny day and fence wires seem darker during the overcast day. Consequently, non-uniform illumination observably occurs along fence wires).

The apparent displacement of same objects known as parallax (Russ, 2002) and non-uniform illumination observably happen in two images of the same scene taken by a mobile camera from slightly different viewing positions, angles and at different times. The general solution of parallax is to register an input image into the same coordinate system with a reference image. The first step in registering input and reference images is to extract automatically distinctive control points from both input and reference images. In outdoor scene, extracting automatically control points can be very challenging because of background complexity and significant illumination variation. Thus, the first objective of this research is to develop an automated control points extraction method that is robust towards varying illumination.

Registering two flat plane images (i.e., images that do not have depths such as satellite images and remote sensing images) can solve parallax among same objects in both flat plane images by extracting as many as possible control points from the images. However, registering two depth plane images (i.e., images that have depths like images captured a digital camera in which the camera is perpendicular towards the scene) cannot solve parallax that occurs in both depth plane images. After registering two depth plane images by choosing several control points from both images, pixels on and close to control points have zero and small gaps referred to as disparity and pixels far from control points have large disparities. This phenomena leads to an investigation of occlusion regions in both registered input depth and reference depth images because these occlusion regions can become potential presence and/or absence of objects in both images. Therefore, the second objective of the research is to develop an automated illumination invariant change detection algorithm in order to extract occlusion regions from depth plane images by investigating stereo correspondence algorithms. Occlusion regions extracted from two depth plane images could still contain potential significant and unimportant changes. Thus, the third objective of the research is to develop intelligent decision-making systems that can decide which objects in the occlusion regions belong to significant or unimportant changes.

Prior to detect breaches in the integrity of and attached objects in front of fence wires, edges of fence wires must be detected first. Detecting edges of outdoor fence wires is a complex task because of their tiny and thin sizes and non-uniform illumination that occurs across fence wires. Thus, the fourth objective of the research is to develop an illumination invariant edge detection approach including a particular change detection algorithm to these particular breaches and attached objects detection.

This dissertation does not present a completely proven method of change detection, since it has not been tested in a real patrol robot. The patrol robot (i.e., a quad bike transformed into a mobile robot) is still under construction and testing. To simulate the movement of a patrol robot, a camera moved manually is used in this research. Thus, a discussion of the patrol robot was outside the scope of this research.

A single camera is used in this research instead of a stereo camera pair. In image change detection, a reference image is needed as an early knowledge of the scene. Multiple input images of the same scene will be somehow compared with the reference image in detecting significant changes while rejecting unimportant changes. The use of a stereo camera pair is not effective since the stereo camera pair produces two stereo images every time in capturing an input image. Moreover, since the research deals with multiple outdoor images of the same scene containing fence wires, natural objects in outdoor scenes are often less of features especially fence wires. As a consequence, stereo correspondence algorithms may fail to produce high-quality disparity maps from the stereo outdoor images.

Due to complexity of problems that have to be tackled in this research, the movement of a mobile camera used in this research was restricted into only three degrees of freedom (3DOF) as other motions can be mechanically overcome. The 3DOF are: (1) movement in left and right directions (X axis), (2) movement in forward and backward directions (Y axis) and (3) rotation in clockwise and anticlockwise directions ($\Delta\theta_z$ axis). The distance of movement in X and Y axes was restricted in a range of ± 20 cm, and the rotation of $\Delta\theta_z$ axis was limited in a range of ± 15 degrees. These movement and rotation restrictions were chosen in simulating navigation errors caused by patrol robot sensors such GPS, compass and accelerometer accuracies.

The presented change detection method was only tested during the day since visible images were used in this research due to time and funding limitations. Therefore, IR images and thermal images were considered outside the scope of this research. In addition, since there are several kinds of wire fences such as taut wire fences, chain-link mesh fences, welded mesh fences and galvanized razor wire fences, an outdoor scene that contains a chain mesh fence was chosen as a sample in this research.

1.3 Structure of the Thesis

Chapter 2 reviews literature to image registration algorithms, stereo correspondence algorithms and current change detection algorithms for both static and mobile cameras. Current gaps in the literature are also summarised, and research aims are specifically stated.

Chapter 3 describes an initial study result of an indoor change detection method. An automated indoor change detection method for multiple indoor images of the same scene acquired by a mobile camera from slightly diverse viewing positions and angles, and at slightly different times is presented in this chapter.

Chapter 4 discusses a template matching-based approach used in this research in order to extract automatically control points from both reference and input images. A current local feature extraction method known as the Scale Invariant Feature Transform (SIFT) operator (Lowe, 2004) is also described. Additionally an explanation of how reference and input images used within this research were captured is described. Moreover, experimental results and concluding remarks of performing the template matching-based approach are presented in this chapter.

Chapter 5 presents an automatic image registration method deployed within this research and how regions of interest are automatically extracted from both reference and registered input images. Experimental results and concluding remarks of image registration are described in this chapter.

Chapter 6 describes how to generate confidence map and occlusion map images. A stereo correspondence algorithm known as the Zitnick and Kanade algorithm (Zitnick and Kanade, 2000) is also briefly discussed. Experimental results and concluding remarks are also described.

Chapter 7 describes an approach used within this research in detecting objects presence and absence behind fence wires. An intelligent decision-making system is applied in this approach in order to decide which objects in occlusion map images that belong to significant or unimportant changes. The decision-making system consists of three sub-systems: crisp, fuzzy and template-subtraction decision-making sub-systems. Output of this approach is a first changed mask.

Chapter 8 presents an illumination invariant edge detection approach used to extract and enhance edges of fence wires. The Sobel detector and an adaptive thresholding technique are also described.

Chapter 9 discusses an approach for detecting breaches in the integrity of and attached objects in front of fence wires. This approach uses the edge detector discussed in Chapter 8 as a core algorithm in extracting edges of fence wires. A specific change detection algorithm is also developed for the purpose of detecting breaches and attached objects. Another hybrid decision-making system is applied in this approach. Output of this approach is a second changed mask.

Chapter 10 presents experimental results and discussion of this research. A fuzzy inference system used to calculate latest possible percentages of significant changes is also described. Estimated locations of and possible percentages of significant changes are displayed on registered input images. Objective, subjective and computational complexity evaluations are provided in this chapter in examining the reliability of the outdoor change detection method. Moreover, the effects of camera movement and rotation restrictions, and of object shadows are also discussed.

Chapter 11 provides conclusions that can be drawn from this research and contributions of the research. Suggestions on possible future work to extend this field of research are also presented.

CHAPTER 2

LITERATURE REVIEW

Image change detection is of widespread interest due to a large number of applications in various disciplines such as video surveillance (Collins *et al.*, 2000; Stauffer and Grimson, 2000; Wren *et al.*, 1997), remote sensing (Bruzzone and Prieto, 2002; Collins and Woodcock, 1996; Huertas and Nevatia, 2000), medical diagnosis and treatment (Bosc *et al.*, 2003; Rey *et al.*, 2002), underwater sensing (Edgington *et al.*, 2003; Lebart *et al.*, 2000) and driver assistance systems (Fang *et al.*, 2003). Regardless of the variety of applications, researchers in image change detection utilize many common pre- and post-processing steps (e.g., image registration, and morphological operations) and change detection algorithms (e.g., image differencing, significance and hypothesis tests, predictive models and the shading model). In this Chapter 2, any potential steps that can be employed in solving problems as mentioned in Section 1.2 are reviewed including image registration, change detection algorithms for both static and mobile cameras, and stereo correspondence algorithms. Current gaps in the literature and research aims of the research are also highlighted at the end of this chapter.

2.1 Image Registration

Two images of the same scene acquired from sightly different positions and angles create the displacement of same objects in the two images known as parallax (Russ, 2002). Parallax is a spatial problem. It can be theoretically solved by registering the two images.

Image registration is one of the fundamental tasks in the image processing and analysis. It is used to match two or more images of the same scene acquired at different times, from different positions, and/or from different sensors. Image registration is the process of spatially aligning multiple images of a same scene (Goshtasby, 2005).

Image registration is widely applied in remote sensing, medical imaging, video processing, and computer vision. According to Zitova and Flusser (2003), the general applications of image registration can be divided into four domains based on how the image is acquired.

1. Different viewpoints (multiview analysis)
   Image registration is performed in images captured from different viewpoints. These kinds of images can be utilized in image mosaicking in order to produce a better overview look image (Su *et al.*, 2004).

2. Different times (multitemporal analysis)
   Image registration is applied to images acquired at different times. Clouds and solar illumination can vary at different times, and these effects need to be taken into consideration during the registration. Multitemporal analysis is used in monitoring land cover change (Dai *et al.*, 1996).

3. Different sensors (multimodal analysis)
   Image registration is performed to images taken from different sensors in order to have images with better quality in terms of spectral and spatial resolutions referred to as data fusion in remote sensing. For example, radar and optical satellite images are merged in improving the effects of clouds and solar illumination variation (Hong and Schowengerdt, 2005; Lampropoulos *et al.*, 2003).

4. Template Analysis
   Corresponding between newly sensed data and a previously developed template (dataset) is compared (Amit and Kong, 1996). In remote sensing, aerial and satellite images can be compared with land maps. It is particularly useful in land cover/land use mapping.

The following steps are usually performed to register multiple images of a same scene captured at different times, from different viewpoints, and/or different sensors (Goshtasby, 2005; Zitova and Flusser, 2003).

1. Pre-processing

   Input images to be registered often have degradation because of sensor nonlinearities, motion blur, noise, and scale differences. The images need to restore firstly by removing noise, deblurring, or rescaling.

2. Feature detection

   A set of features are detected manually or automatically from both input and reference images. There are two main methods for the feature detection: area-based and feature-based algorithms. In area-based methods, a small window of points in the reference image is statistically compared with windows of the same size in the sensed image. The comparison uses a similarity metric, which measures the similarity between two given windows. In feature-based algorithms, the image is represented in a compact form by a set of features. The features are invariant to the scaling, rotation, and gray level modification. The common features are edges, regions, lines, line endings, line intersections, or region centroid.

3. Feature matching

   The correspondences between the features detected in both input and reference images are established. Various feature descriptors and similarity measures along with spatial relationships among the features are used for that purpose.

4. Determination of a transformation function

   Knowing the coordinates of a set of corresponding features in both input and reference images, the transformation parameters in a mapping function for registration can be determined. In order to define the mapping function, a priori information is needed regarding degradations. If there is no a priori information, mapping functions must be flexible so they can be suitable to all possible combinations of degradations.

5. Image resampling and transformation

   After the transformation, the registered image pixel coordinates are not integers anymore. The corresponding integer-valued pixel intensity is computed by an appropriate interpolation technique.

To carry out those steps above automatically, several methods have been proposed and are divided into four groups (Reddy and Chatterji, 1996); (a) methods that directly use image pixel values, (b) methods that operate in the frequency domain, (3) methods that use low-level features such as edges and corners, and (4) methods that use high-level features such as identified objects or features. The followings are further explanations of the methods.

1. Wavelet-Modulus Maxima Method

   Fonesca and Costa (1997) use image pixel values. The probable control points are detected from the local modulus maxima of the wavelet transform applied to the input and reference images after performing the wavelet decomposition up to two levels. The correlation coefficient is used as a similarity measure and only the best pair-wise fitting among all pairs of feature points are taken as actual control points. A polynomial transform which can take care of translation and rotational errors is then used to model the deformation between the images and their parameters are estimated in a coarse to fine manner. The refinement matching is achieved using the warped image and the set of feature points detected in the reference image. After processing all levels, the final parameters are determine and used to warp the original input image.

2. Fast Fourier Transform (FFT) Method

   Reddy and Chatterji (1996) and Keller and Averbuch (2002) work in a frequency domain approach in which it doesn't use any control points, instead the FFT ratio is computed. The displacement between two given images can be determined by computing the ratio using the following formula:

   $$F_1.\mathit{conj}(F_2) \,/\, \mathit{abs}\ F_1.F_2 \qquad (2.1)$$

   in where F_1 is the Fourier transform of image 1, *conj* is the complex conjugate, F_2 is the Fourier transform of image 2 and *abs* is the absolute value. The inverse of this ratio results as an impulse like function. This is approximately zero everywhere except at the displacement (it determines the translation error between the images). Converting these images from rectangular coordinates to log-polar coordinates and by calculating the similar ratio, we can represent rotation and scaling errors also as shifts. These three parameters are used to establish the mathematical model and the image is geometrically rectified with respect to the reference image.

3. Morphological Pyramid Image Registration Method

Hu and Acton (2000) use the low level shape features to determine the global affine transformation model along with the radiometric changes between the images. The multiresolution images are represented by a Morphological Pyramid (MP) as the MP has capability to eliminate details and to maintain shape features. The MP of an image is consecutively structured by morphological filtering and sub-sampling:

$$I_L = \lfloor (I_{(L-1)} \circ K) \bullet K \rfloor_{\downarrow d},\quad L = 0,1,2,\ldots,n \tag{2.2}$$

where L is the pyramid level, I_0 is the original image, $[\,]_{\downarrow d}$ represents a down sampling by a factor of in each spatial dimension along rows and columns, $(I \circ K)$ is the morphological opening of the image I with structuring element K, and $(I \bullet K)$ represents the morphological closing. The Levenberg Marquardt non-linear optimization algorithm is employed to estimate the matching parameters of translation, rotation and scaling errors up to sub-pixel accuracy. In this approach, intensity mapping function is integrated into geometric mapping system.

4. Registration using Genetic Algorithm (GA) Method

Genetic algorithm is an iterative procedure that maintains a population of candidate solutions encoded in the form of chromosome strings (Goldberg 1989). Genetic algorithm is used to efficiently explore the huge solution space required by the image registration to sub-pixel accuracy. Highest similarity between the input and reference images indicates the proper registered images. The similarity can be achieved by properly identifying the correct transformation procedure. Genetic algorithm adaptively explore the search solution search in a hyper dimension fashion, therefore that genetic algorithm can improve computational efficiency.

Chalermwat and El-Chazawi (1999) propose that each chromosome is of length 32 bits, allocates 12 bits for rotation, 10 bits for translation in x-direction and 10 more bits for translation in y-direction. Each field is a signed magnitude binary number. A precision factor is used to improve the accuracy. Evaluate the fitness function for each solution in the population to see if the termination criteria for optimality are met. They used a weighted roulette wheel sampling to reproduce strings of the next generation in proportion to their fitness. Evaluate the fitness of each new individual. They used affine transformation and bilinear interpolation. Population size and number of generations were limited to 150, registration accuracy observed as less than a pixel.

Images used in the previous methods were flat plane images. As images used in this research are depth plane images acquired by a mobile camera, intensity values of pixels and shapes of same objects in the depth plane images change significantly because of camera motion, illumination variation and background clutter. Thus, developing an automated image registration that can overcome changes in shapes of same objects and intensity values of pixels is crucial in this research in order to minimise parallax in the depth plane images.

2.2 Stereo Correspondence Algorithms

No matter how image registration is undertaken, the displacement of the same points in the reference image and the registered input image presents disparity (range). The disparity between a specific object (target) and other objects behind and in front of the target will be different. All object disparities are summarized in a disparity map. A confidence map can often be produced based on information of the disparity map that depicts the confidences of match values which are correct.

In order to obtain the disparity map and/or the confidence map, several stereo correspondence algorithms have been developed. They are divided into two algorithmic paradigms, referred as feature- and as area-based approaches (Dhond and Aggarwal, 1989). In the following paragraphs, two classes of feature-based approaches that have received recent attention are discussed: hierarchical feature matching and segmentation matching.

Venkateswar and Chellappa (1995) presented an algorithm that improves stereo and motion matching. In their research, they utilized four types of features: lines, vertices, edges and edge-rings. Matching commenced at the highest level of the hierarchy (edge-rings) and continued to the lowest (lines). The feature-based hierarchical framework allowed coarse, reliable features to provide support for matching finer, less reliable features, and it reduced the computational complexity of matching by reducing the search space for finer levels of features. The feature hierarchy was built from the bottom up by firstly extracting edges of objects based on structural (i.e., connectivity) and perceptual (i.e., parallel, collinear and proximate) relationships. Incompatibility relations (e.g., intersects, overlaps and touches) were utilized to enforce consistent feature groupings. All potential features in the hierarchy were stored as hypotheses in a relational graph. Inconsistent groupings were cut from the graph by a truth maintenance system (TMS). Feature matching was then performed between the relational graphs of the stereo images, commencing with edge-rings and continuing to lines. Once a higher-level feature match had been confirmed, the component features were no longer included in the search for other lower-level matches, since a feature could not belong to more than one group. This reduced the search space significantly at each level of the hierarchy.

Birchfield and Tomasi (1999) segmented stereo images into small planar patches for which correspondence is then determined. As with most feature-based methods, this reduces the match sensitivity to depth discontinuities. However, these planes are likely to be slanted rather than fronto-parallel (i.e., directly facing the cameras), so the relationships between segments in the two images are modelled by six parameter affine transformations, such that

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = A \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + d \tag{2.3}$$

$$A = \begin{bmatrix} 1+d_{xx} & d_{xy} \\ d_{yx} & 1+d_{yy} \end{bmatrix} \quad and \quad d = \begin{bmatrix} d_x \\ d_y \end{bmatrix} \tag{2.4}$$

where (x_1, y_2) and (x_2, y_2) are the coordinates of corresponding points in the left and right images respectively, and the vector *d* defines the translation of a segment between frames. In the case of rectified stereo images, d_{yx}, d_{yy} and d_y values are 0. In the affine transformation, the matrix *A* can be used to define the in-plane rotation, scale and shear transformations between frames. The parameters are computed directly from spatio-temporal intensity gradients. For epipolar-rectified imagery, only the horizontal parameters are computed. Segmentation and affine parameter estimation are computed iteratively, and patches with similar affine parameters are merged after each iteration. The segmentation algorithm used is based on the multiway cut algorithm of Boykov *et al.* (1998). Unlike most feature-based methods, dense disparities are explicitly defined for this segmentation-based method by planar transformations. However, this approach is also sensitive to the quality of the original segmentation.

In the following paragraphs, we review a few classes of area-based stereo approaches. Zitnick and Kanade (2000) presented a cooperative stereo algorithm using global constraints to find a dense depth map. A three-dimensional array of match values is constructed in disparity space; each element of the array corresponds to a pixel in the reference image and a disparity, relative to another image. An update function of match values is constructed for use with real images. The update function generates continuous and unique values by diffusing support among neighboring match values and by inhibiting values along similar lines of sight. Initial match values, possibly obtained by pixel-wise correlation, are used to retain details during each iteration. After the match values have converged, occluded areas are explicitly identified. In other words, they combined diffusion of the support region with inhibition of support for pixels along similar lines of sight. This ensured that the diffusion process did not violate the uniqueness constraint. However, the computation of the update function (correlation measure) led the cooperative stereo algorithm to be a computationally expensive process, since extensive search was required in configuration space.

Okutomi and Kanade (1992) addressed the issues of window size selection and presents a statistically sound technique which minimises the uncertainty in the disparity estimate at each pixel of the depth map. The authors make the observation that larger matching windows provide better disambiguational ability and less accuracy, and attempt to address this problem by optimally selecting a window size in a dynamic fashion as the matching process proceeds. This demonstrates the use of a principled technique for determining an otherwise arbitrary parameter. Although, in practice, it would seem that either a hierarchical matching technique or a deformable surface model would have the same effect.

2.3 Change Detection Algorithms for Static Cameras

The goal of a change detection algorithm is to categorize image pixels into two sets: changed and unchanged. In common cases, the image differences caused by motion objects relative to background, appearance or disappearance of objects and shape, colour and texture changes of objects are considered to be significant changes. Ambient and sensor noise, camera motion, illumination variation and registration error, which cause the image differences, are considered as unimportant changes. In the following paragraphs, a systematic survey on these approaches is presented.

2.3.1 Predictive Model

The idea of this approach is to formulate the gray value intensity in a given region as a polynomial function of the pixel coordinates. A representative of this approach is the quadratic picture function model proposed by Hsu (1984). He modelled an image as a mosaic of blocks where the intensity value was formulated as a second-order bivariate polynomial function of the pixel coordinates. Change detection was carried out by comparing corresponding block pairs in two images. If two blocks can be least-square fit by a same group of polynomial coefficients, then no change was detected between the two blocks. The alternative decision would be drawn if they were best fit by different polynomial coefficients. The major weak point with this approach was the assumption that image intensity could be modelled as quadratic function was often violated in real scenarios. In addition, the residuals from the polynomial fit might not be Gaussian distributed either. Therefore, the accuracy of the likelihood test was unreliable.

2.3.2 Hypothesis Testing

In this approach, whether a pixel was changed or unchanged was determined by choosing the hypothesis that best matches the observation and the prior knowledge. The significance test developed by Aach and Kaup (1995) was a typical hypothesis testing approach. In this approach, the statistics of noise was utilized to test whether the observed image difference was caused solely by noise. The null hypothesis in the test was that under the condition of no change, the image difference could be modelled as a random variable that had a zero-mean Gaussian distribution with a known variance. The test of this hypothesis was carried out at each local region which was a spatial window centred at the testing pixel. The testing variable was defined as the local sum of squared difference of the pixel intensity normalized by the noise variance. This variable under the null hypothesis had a χ^2 distribution with the degrees of freedom equal to the number of pixels inside the local window. Therefore, the decision threshold was determined by specifying a confidence level of the preservation of the null hypothesis. The approach performed change detection heuristically well if the local window size and the confidence level were properly chosen. The weakness of this approach was that the testing was one-sed, meaning that the knowledge of the alternative hypothesis was not utilized at all. As a consequence, this approach lacked the sense of optimality.

2.3.3 The Shading Model

Phong (1975) proposed this approach that intends to exclude illumination variation from significant changes by utilizing the shading model which formulates image intensity based on physical aspects of light reflection. With appropriate assumptions, the gray level intensity of a pixel was approximated by the product of the illumination of a physical surface point and its shading coefficient. This coefficient was determined by a number of factors such as the reflectance of the surface material and angles of striking and reflected lights. If no change undergone the physical structure of an object, the shading coefficient was assumed to be intact. Under such condition, the ratio of pixel intensities in two images became the ratio of illumination from the two corresponding physical locations. Since illumination could be approximated as a constant within regions that were sufficiently small, the pixel intensity ratios remained constant in the testing blocks under the condition of no change.

Based on this rationale, Skifstand and Jain (1989) suggested to test the variances of pixel intensity ratios within two given blocks. If the variance was smaller than a threshold empirically selected, then it was determined that the imaged object surfaces were in the absence of change.

Durucan and Ebrahimi (2001) formulated the change detection from a point of view of linear dependence test. They formulated the hypothesis of no change as linear dependence between vectors of corresponding pixel intensities. The test was carried out by thresholding the determinants of

Wrongskian matrices that represented the linear dependence of the given vectors.

Both Skifstad and Jain's and Durucan and Ebrahimi approaches were centred around the shading model, in which illumination variation was dealt with reasonably well. However, the noise effects were not considered in these models. As a consequence, the thresholds utilized in these tests were chosen in an ad hoc manner.

2.3.4 Background Modelling

In the context of surveillance applications, change detection is closely related to the well-studied problem of background modelling. The goal is to determine which pixels belong to the background or foreground (changed pixels). A large amount of frame-rate video data is available, and the images between which it is desired to detect changes are spaced apart by seconds. The entire image sequence is used as the basis for making decisions about change, as opposed to a single image pair.

Haritaoglu *et al.* (1998) used grayscale information in their method. First of all, a background model was generated with number of frames in which each pixel was described with three values as the following: minimal intensity (M), maximal intensity (N) and maximal difference value of two successive frames. The difference images were calculated with current image and both the M image and N image. These were used for the classification in which the foreground pixels were given if the difference values were greater than the values of the maximal interframe difference. After the binarization some post processing steps were also needed for elimination of noises. Furthermore, the classified background pixels were then used for updating the background model for considering sudden environmental changes.

Francois and Medioni (1999) assumed that in the background only very slow global changes can be occurred and further the colour values of each pixel build a sphere cluster in the Red, Green and Blue (RGB) colour space. With these assumptions a background model as a Gaussian distribution was generated by considering the mean value and standard deviation for each pixel.

The Hue, Saturation and Value (HSV) colour space was used instead of the RGB. The current image was subtracted from the mean value model and the resulted difference values of each pixel gave the information of classifying to either foreground or background regarding to the standard deviation model. Moreover, an update of the background model was also given.

Similar to Francois's assumption, McKenna *et al.* (2000) modelled the background with mean value and standard deviation. However, their system considered two parameters: the normalized RGB colour space and the edge. For each channel the models were generated. For a number of frames, two models were calculated in order to separate colour and edge. For both issues the current image was converted to the adequate form such as edge image and RGB image which were used apart for the further classification. At least a combination of both classification results gave the final segmentation mask.

Cavallaro and Ebrahimi (2002) proposed another approach. For each channel of used YCbCr colour space an image differencing with background and current images was applied. With each preliminary result an edge detection algorithm was utilized using the sobel algorithm. Then all three sub results were fused together which still occurred problems since the detected edges were not really connected as a whole contour. Therefore, a post processing step was needed such as the morphological operations to get the resulted mask.

Hong and Woo (2003) also modelled the background, but this time both well-known RGB and normalized RGB colour were applied. As mentioned in previous methods the mean and standard deviation were used again and these were calculated over each colour component. Each colour space had its own classification part in which the current image was converted first in each colour space. Within each colour space the pixel could be classified in four categories: background, foreground with shadow, background with shadow and foreground.

Shen (2004) used the well-known RGB colour space and the system could be represented in two sections. One of them was the block for generation of fuzzy classification and the other one was the block for elimination of falsely detected segmentation regions. The fuzzy classification was applied to take into account the mobility of pixels precisely instead of the so called binary classification. Thus, in the fuzzy block a difference image was generated for each RGB colour space component. For every channel's result a corresponding threshold was determined by use of unimodal thresholding method for considering the fuzzy set of mobile pixels. Then these thresholds availed to generate fuzzy images which at least were combined to one final fuzzy image. Subsequently, a preliminary mask was achieved by thresholding which described all detected mobile pixels in all appearances.

To overcome the problems of illumination changes and since there was no sudden adaptive update of the background a combination of temporal information and the mentioned above fuzzy colour classification was given. The temporal information was achieved by the OR operation of the image differencing of successive frames and the last resulted mask. This output was combined with the preliminary mask of the fuzzy classification block.

Those algorithms presented in this section used multiple images of the same scene taken by a static camera as inputs. They detected significant changes while rejecting unimportant changes mainly caused by illumination variation. Since this research deals with multiple images of the scene captured by a mobile camera, the complexity of change detection increases significantly in these kinds of multiple images caused mainly by camera motion, background clutter and illumination variation. Thus, new change detection methods are researched and developed for these kinds of multiple images.

2.4 Change Detection Algorithms for Mobile Cameras

A mobile single vision is used in this project in order to acquire multiple images of a scene of interest. In the following paragraphs, a review of using mobile cameras in several different real-time applications is described.

Environment equipped with multiple cameras was used for counting the number of different people walking through such environment (Kettnaker and Zabih, 1999). The researchers used four different camera views. The proposed solution was to analyse the content of multiple camera streams together by considering floor topology, global consistency constraints, and temporal dependency of successive camera appearances of people walking along corridors. Although multiple cameras could be applied in a video-surveillance system in order to patrol such large protected areas, the system could be relatively too expensive in term of hardware needs and computation time.

A real-time surveillance system equipped with a single camera on a pan/tilt platform was presented to track multiple moving targets within the camera's field of regard (Benabdelkader *et al.,* 2000). The aim of this design was to maintain motion trajectory information for as many moving targets as possible. The solution consisted of two procedures. The first procedure handled the target scheduling problem within a queuing theory framework. The other handled low-level localized detecting and tracking target based on estimation and feedback control. However, the solution was computationally heavy and did not allow a real-time process, which was necessary for a video-surveillance system.

Oberti *et al.* (2002) presented a video-surveillance system based on a pan/tilt mobile camera. Their method began by creating a panoramic multi-layer background image. The background image and the camera position when capturing an input image were used to perform change detection. While the system could detect a new object in a video sequence from a non-static camera, the system depended heavily on the current position of the camera when capturing an input image. When this position of the camera was unknown, the system failed to detect changes in the input image

Primdahl *et al.* (2005) described an approach that analyze videos acquired from a moving vehicle making repeated passes through a specific, well-defined corridor. The objective of their approach was to detect stationary objects that appear in the scene along the established route. Their approach began with pairing images of different videos. For each frame pair, they introduced regions of interest. Although the new objects were successfully identified out of the videos captured from different camera locations, this only occured because the authors manually selected four points on the regions of interest that contained the new objects. In other words, the authors explicitly identified the locations of the new objects.

2.5 Summary of Literature

The research deals with detecting regions of change in multiple images of the same scene acquired by a mobile camera from slightly different viewing positions, angles and at different times while rejecting unimportant changes caused mainly by camera motion, illumination variation and sensor noise. Complexity of the research significantly increases when detecting regions of change (i.e., appearing and disappearing of objects behind fence wires, breaches in the integrity of fence wires and attached objects in front of fence wires) is performed in multiple outdoor images of the same scene containing fence wires (i.e., a chain-link mesh fence) captured by a mobile camera, since background clutter, tiny sizes of fence wires and non-uniform illumination that occurs along fence wires during the day have to put into consideration. Therefore, the research is about investigating and developing a new automated change detection method for these kinds of multiple outdoor images. Referring to knowledge of the author and the literature, the research is the first investigation in term of change detection methods for these kinds of multiple outdoor images. There are several problems that arise from these kinds of multiple outdoor images: displacement of same objects known as parallax (Russ, 2002), significant illumination variation during the day, background clutter of outdoor scenes, tiny sizes of fence wires and non-uniform illumination on surfaces of fence wires. The new automated change detection method has to be able to overcome these problems.

Parallax could be solved by registering two images of the same scene (Goshtasby, 2005; Zitova and Flusser, 2003). However, parallax can only be solved successfully if the two images are flat plane images like remote sensing images captured by a satellite. In this research, images investigated are depth plane images. In depth plane images, no matter how image registration is undertaken, the displacement of the same points in two images presents the difference (Russ, 2002).

Prior to perform image registration two images, correspondence points must be extracted automatically from both images. Area-based and feature-based methods are available in literature in extracting control points from two images captured by static cameras (Orduyilmaz, 2006). However, this research uses indoor and outdoor images taken by a mobile camera. As a result, size, position and intensity values of same objects in the indoor and outdoor images change significantly. This research will investigate a new approach in extracting control points from multiple indoor and outdoor images of the same scene captured by a mobile camera.

Occlusion regions provided by stereo correspondence algorithms (Zitnick and Kanade, 2000; Venkateswar and Chellappa, 1995; Birchfield and Tomasi, 1999) could contain potential significant changes such as appearing new objects and disappearing previous objects in the scene. As mentioned above, registering two depth plane images presents the gap (disparity) in between the same points in the two images. This phenomenon will be investigated deeply in order to detect the occlusion regions.

Occlusion regions could contain potential significant changes as well unimportant changes. A decision-making system like a fuzzy inference system will be developed in this research in order to decide which objects in the occlusion regions belong to significant or unimportant change.

Detecting edges of fence wires in multiple outdoor images of the same scene containing fence wires is a challenging task. Available edge detectors, such as the Sobel, Prewitt, Roberts (Wang *et al.*, 2003), and Canny (Peihua, 2007) operators, can be used to detect edges of fence wires. Edged images produced by these edge detectors are still in gray images. For a further process like boundary detecting, edged images have to be converted to binary images. These edge detectors utilize global threshold values in converting gray images to binary images. Since non-uniform illumination observably occurs along fence wires during the day, using global threshold values may not be the best solution. Thus, an adaptive thresholding technique will be researched and utilized in this research in order to overcome non-uniform illumination that occurs on fence wires.

To summarise this section, there are several aims that have to be achieved in this research. Firstly, the research develops a simple indoor automated change detection method for multiple indoor images of the same scene acquired by a mobile camera from slightly different viewing positions, angles and at slightly different times. The reason of performing an experiment in indoor environment is to get an initial knowledge of how a change detection method for multiple images of the same scene captured by a mobile camera overcomes parallax in these kinds of multiple indoor images. Secondly, based on the result of indoor change detection, the research develops an automated change detection method for multiple outdoor images of the same scene containing a chain-link mesh fence taken by a mobile camera from slightly dissimilar viewing positions, angles and at different times. Problems mentioned in Section 1.2, such as extracting control points for image registration in a natural outdoor scene, minimizing parallax by performing image registration, generating confidence map and occlusion map images and developing intelligent decision-making systems, are tackled in this outdoor change detection method. Finally, multiple outdoor images of the same scene containing a chain-link mesh fence captured by a mobile camera from three different outdoor scenes are fed into the outdoor change detection method in order to examine the reliability of the method. Objective, subjective and computational complexity evaluations are used to assess the consistency of the outdoor change detection method.

CHAPTER 3

INDOOR CHANGE DETECTION METHOD

Chapter 3 presents a simple automated change detection method for multiple indoor images of the same scene taken by a mobile camera from slightly different viewing positions and angles. This chapter also explains how reference and input indoor images used in this chapter were captured by a mobile camera and presents experimental results and discussion.

3.1 Background

Displacement of same objects (parallax) observably appears in two images of the same scene taken by a mobile camera from slightly different viewing positions and angles. To minimize parallax in the two images, a current image (i.e., an input image) is often registered into a same coordinate system with a previous image (i.e., a reference image). In the case of flat plane images, registering two images of the same scene are performed by extracting as many as possible correspondence points from both flat plane images. Since this research deals with depth plane images captured by a mobile camera, extracting control points from depth plane images is a challenging task because of the sizes of same objects in the depth plane images will change significantly.

To get a preliminary knowledge of how to minimize parallax in multiple depth plane images of the same scene captured by a mobile camera, an initial study was performed in the early experiments of this research and an indoor environment was designed. Illumination variation was not included into a consideration yet. Objects such as pictures, a table, a wall, a camera, a pencil and an eraser were utilized in this indoor experiment. An indoor change detection method for these kinds of multiple indoor images is described in the next section.

3.2 Algorithm Overview

The indoor change detection method consists of three major steps: (1) automatic image registration (AIR), (2) temporal differencing and (3) unimportant changes removal (UCHR). Fig. 3.1 shows flow chart of the change detection method.

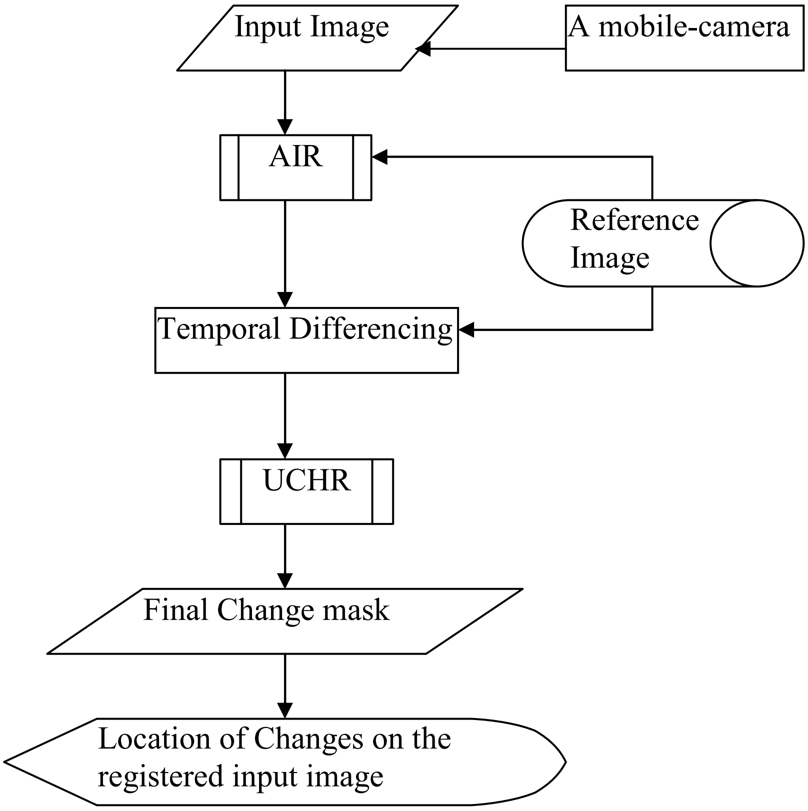


Fig. 3.1 Flow chart of the indoor change detection method

3.2.1 Automatic Image Registration

The Scale Invariant Feature Transform (SIFT) operator [Lowe, 2004] was used to automatically extract and match local features from both reference and input images in this study. The local features were extracted by the SIFT operator from three templates which were chosen and cropped manually off-line in advance from the reference image. The local features provided by the SIFT operator were invariant to rotation, scaling, and slightly tolerant of small changes in illumination.

The following were steps established in this study in order to perform automatic image registration of an input image:

1. Read the colour reference image and the current colour input image.
2. Convert colour reference and input images into gray reference and input images.
3. Read three gray template images.
4. Extract three control point locations from the reference image
   4.1. Perform the SIFT operator towards the template 1 and the reference image,
   4.2. Extract only matched keypoint locations in the reference image,
   4.3. Calculate an average value from the extracting matched keypoint locations and use the average value as the first control point location,
   4.4. Repeat steps 4.1 – 4.3 for template 2 and template 3 for the second and third control point locations.
5. Extract three control point locations from the input image by repeating step number 4.1 – 4.4.
6. Tune new three control point locations in the input image by using the cross-correlation method based on information of the original control point locations in the input image (control point locations provided in step number 5), the three control point locations in the reference image, the input image, and the reference image. $x\min = 1$
7. Estimate parameters of the linear conformal (similarity) transformation such as scaling, rotational, and translational differences between the reference and input images respectively, based on the three pairs of the correspondence points. The similarity transformation works based on the equation below.

   $$X = s\,x\cos(\theta) - s\,y\sin(\theta) + h \tag{3.1}$$

   $$Y = s\,x\sin(\theta) + s\,y\cos(\theta) + k \tag{3.2}$$

   where *s,* θ, and *(h,k)* are the scaling, rotational, and translational differences between reference and input images.
8. Transform and re-sample the input image based on the estimated linear conformal transformation in order to generate the registered input image.

3.2.2 Temporal Differencing

In order to produce a preliminary change mask, the temporal difference algorithm [Chang *et al.*, 2005] was firstly used to subtract the reference image and the registered input image. If the reference image is assumed as R(x,y) and the registered input image is I1(x,y), then the subtracted image is

$$D(x,y)=\left|R(x,y)-I1(x,y)\right| \tag{3.3}$$

As a result of this subtracting, an overlapping region appeared on the subtracted image. Secondly, the overlapping area, referred to region of interest, was automatically cropped based on the registered input image information.

Image thresholding was finally performed to the region of interest. The result of this thresholding was a preliminary change mask (B(x,y)) created according to the following decision rule:

$$B(x,y)=\begin{cases} 1 \; if \; RoiD(x,y) \rangle \tau, \; change\,occured \\ 0 \; if \; RoiD(x,y) \leq \tau, \; no\,change \end{cases} \tag{3.4}$$

where τ is a global threshold value and RoiD(x,y) is region of interest of the subtracted image. To accelerate searching for significant changes, searching was only applied to the region of interest.

3.2.3 Unimportant Changes Removal

In the sub-section 3.2.2, simple differencing followed by thresholding has the advantage of low computational cost. However, the result is sensitive to noise and shadow. In addition, misalignment of the input image also causes unimportant changes appearing in the preliminary change mask.

The misalignment happened because control points extracted from the three templates could not cover the whole area of the reference and input images. Hence, the following operations were designed to remove unimportant changes from the preliminary change mask.

Clearing Border Objects

In this operation, objects that touch the border of the preliminary changed mask are removed. By using morphological reconstruction, in which the preliminary changed mask is used as the mask, and the marker image, f_m, is defined as

$$f_m(x,y)=\begin{cases} B(x,y) & if\,(x,y)\; is\; not\,on\; the\,border\,of\; B \\ 0 & otherwise \end{cases} \tag{3.5}$$

Erosion Using Vertical and Horizontal Lines

Erosion is a process of shrinking objects in a binary image. The manner and extent of shrinking is controlled by a structuring element. In this case, horizontal and vertical lines were used as the structuring element in order to erode objects in the $f_m(x,y)$.

Removing Small Objects

Next, all connected components (objects) that had fewer than 100 pixels were removed from a binary image ($f_m(x,y)$). In this real life setting, 100 pixels represent an object with area of 480 square millimetres (size of a standard pencil) when an image is captured from a 230 cm distance. A final changed mask ($f_{m1}(x,y)$)was generated as output of this operation.

$$f_{m1}(x,y) = \begin{cases} f_m(x,y) & if\ (x,y) > 100\ pixels \\ 0 & otherwise \end{cases} \tag{3.6}$$

3.3 Reference and Input Images

In order to test the robustness of the indoor change detection method, several indoor images were used as input in this study including a reference image and 7 input images. Fig. 3.2, below, depicts the reference image used in this experiment.

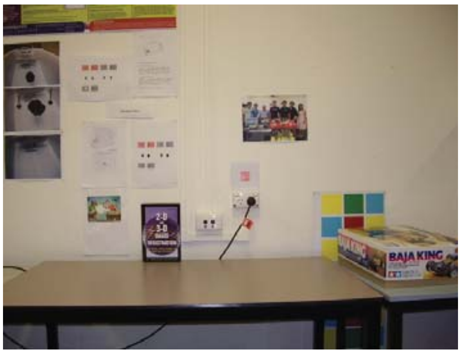


Fig. 3.2 The reference image utilized in this indoor experiment. It was a scene of an indoor environment that consisted of objects such as books, pictures, a wall and a table

It was captured in such a way that the position of a digital camera was perpendicular towards the scene. The distance between the digital camera and the wall of the scene was 230 cm. This current position of the mobile camera was referred as the origin position ((0,0,0) position).

In this study, the movement of the mobile camera was restricted only by three degrees of freedom ((X, Y, $\Delta\theta_z$) axes). The X axis represented translation on left (-) or right(+) sides. The Y axis indicated movement on forward (+) or backward (-) sides (zoom in / zoom out). The $\Delta\theta_z$ axis was rotation on the clockwise (+) or anticlockwise (-) sides. The camera movement in X and Y axes is limited in ± 20 cm and the rotation of the mobile camera in Z axis is limited in ± 15 degrees.

After capturing the reference image, four new objects were added to the scene: a pencil, a small bolt, a conventional camera and a glasses case. The digital camera was then shifted manually to several input camera positions as depicted in Table 3.1 below and Fig. 3.3 depicts top views of camera positions when capturing input images used in the indoor experiment.

Table 3.1 Camera positions when capturing input images

Images	**Positions (X, Y, $\Delta\theta_z$)**	**Positions in Fig. 3.2**
Indoor Input Image 1 (IdII-1)	(0, +20, 0)	1
Indoor Input Image 2 (IdII-2)	(+15, +20, 0)	2
Indoor Input Image 3 (IdII-3)	(+10, +20, 0)	3
Indoor Input Image 4 (IdII-4)	(+5, +20, -5)	4
Indoor Input Image 5 (IdII-5)	(+5, +10, -10)	5
Indoor Input Image 6 (IdII-6)	(-10, +15, +5)	6
Indoor Input Image 7 (IdII-7)	(+10, +15, -15)	7
Indoor Input Image 8 (IdII-8)	(-15, +20, +10)	8

As depicted in Table 3.1 and Fig. 3.3, below, input images were captured by a mobile camera from slightly different viewing positions and angles. Camera positions when capturing input images are referred as input camera positions. These input camera positions were randomly chosen in order to simulate random motions of a patrol robot in the real time application. To capture, for example the seventh indoor input image (IdII-7), the digital camera was shifted manually 10 cm forward (zoom in) and 15 cm to the right. It was then rotated 15 degrees in an anticlockwise direction. This current camera position, (+10, +15, -15), was referred to as the seventh input camera position, from which the reference image was captured. The

■ symbol represents the mobile camera and the red colour symbol (🟥) in Fig. 3.3 represents a position in which a reference image was captured. Figs. 3.4 (a) – (g), below, depict IdII1 – 8, respectively.

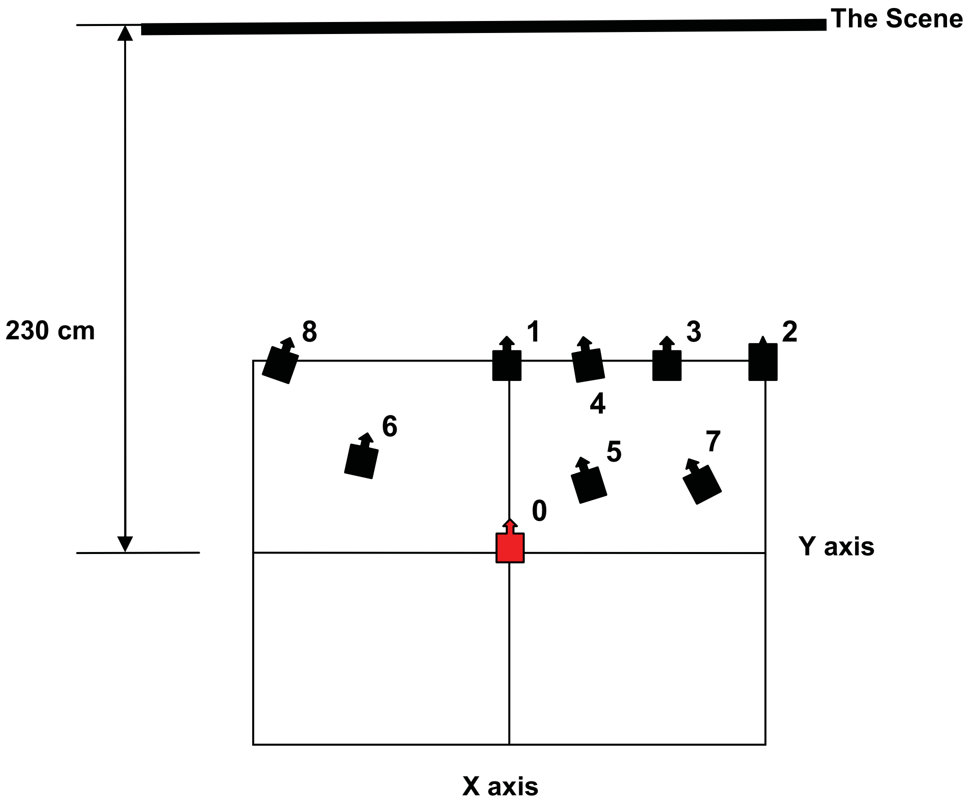
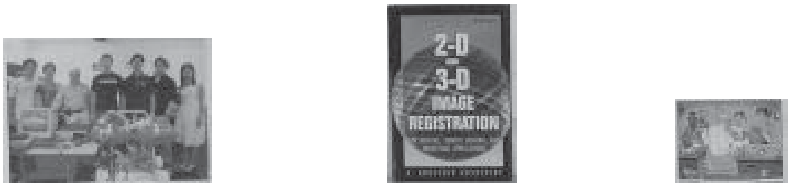


Fig. 3.3 Top views of input camera positions. As seen in Fig. 3.3, the mobile camera has been shifted on the three degrees of freedom ((X, Y, $\Delta\theta_z$) axes)

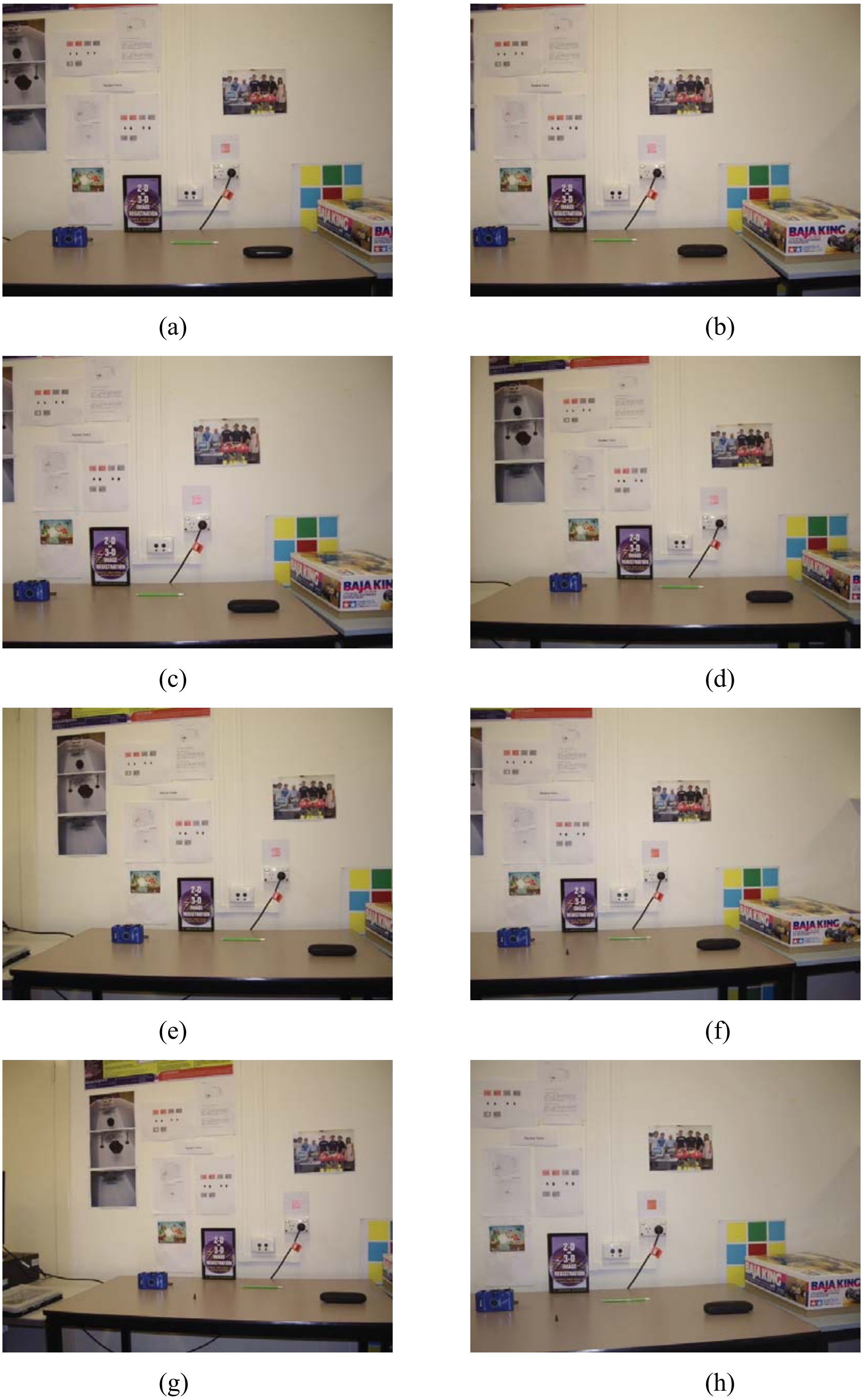
3.4 Experimental Results and Discussion

In order to automatically extract control points from both reference and input images, three templates were chosen and cropped manually in advance from the reference image. These templates are shown in Figs 3.5 (a) – (c), below, respectively.



(a) (b) (c)

Fig. 3.5 Templates extracted in advance from the reference image



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Fig. 3.4 Input images used in this indoor experiment

As seen in Figs. 3.4 (a) – (f), respectively, and Fig. 3.2, the reference image, occluded regions and displacement of same objects (parallax) apparently appear in input images. Occluded regions are composed by potential significant changes and unimportant changes

Any object in the reference image can become a candidate as a template. The main reason for choosing these templates is that they contain many local features. As a sample, the following paragraphs will only describe results of applying the indoor change detection method towards the IdII-7 (see Fig. 3.4 (g)).

The SIFT operator was used in this experiment in order to extract and match local features known as keypoints from these templates and both reference and input images. Fig. 3.6 shows a sample of extracting and matching local features from template 1 and the IdII-7.

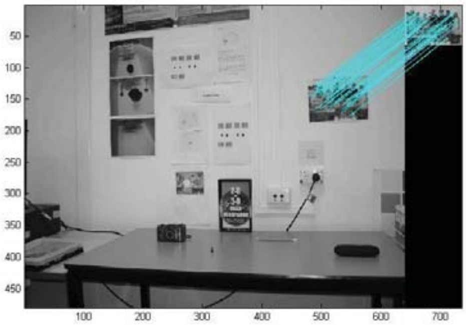
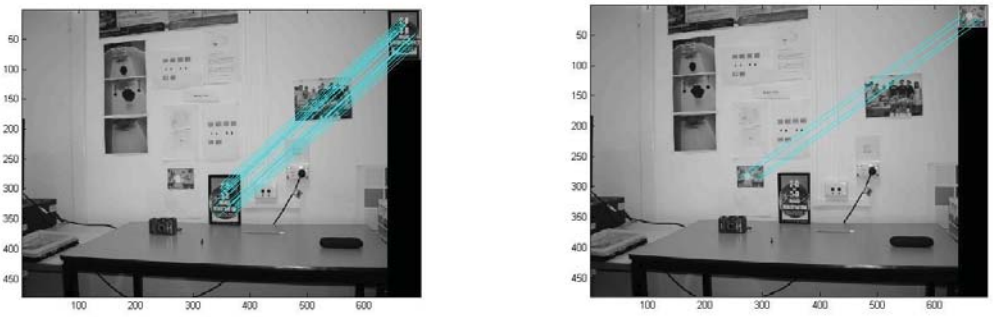


Fig. 3.6 A sample of matched keypoints provided by the SIFT operator

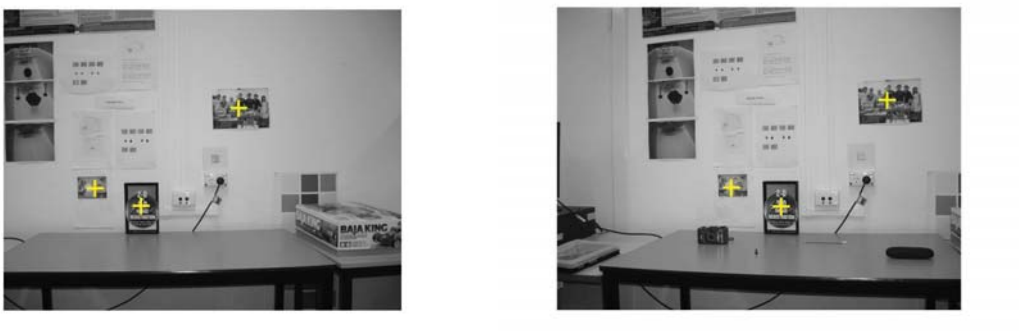
Referring to the image in Fig. 3.6, the SIFT operator extracted 691 keypoints from the IdII-7 and 116 keypoints from the template 1. From these extracted keypoints, 49 were matched. A matched keypoint indicates positions of keypoints found by the SIFT operator matching on the IdII-7 and the template 1 in X and Y axes. By extracting the positions of matched keypoints only in the IdII-7, a control point could be determined by calculating an average value in the X and Y axes.

The same process was also employed for the template 2 and 3. Figs. 3.7 (a) and (b) depict matched keypoints generated by the SIFT operator from the template 2, 3 and the IdII-7. Fig. 3.8 (a) and (b) show control points extracted automatically from the reference image and the IdII-7.



(a) (b)

Fig. 3.7 Matched keypoints provided the SIFT operator from the template 2, 3 and the IdII-7. The SIFT operator extracted 74 keypoints from the template 2, 25 keypoints from the template 3 and 691 keypoints from the IdII-7. Matched keypoints found by the SIFT operator from the template 2 and the IdII-7, and from the template 3 and the IdII-7 were 29 and 4 matches



(a) (b)

Fig. 3.8 Control points extracted automatically from the reference image (a) and the IdII-7 (b)

A tuning process was then performed to increase, to sub-pixel accuracy, the control points on the IdII-7. The tuning process uses the following general procedure. For each control-point pair,

1.1. Extract an 11-by-11 template around the input control point and a 21-by-21 region around the base control point,

1.2. Calculate the normalized cross-correlation (NCC) of the template with the region. The NCC equation is shown below.

$$\gamma(u,v) = \frac{\sum_{x,y}\left[f(x,y) - \overline{f}u,v\right]\left[t(x-u,y-v) - \bar{t}\right]}{\left\{\sum_{x,y}\left[f(x,y) - \overline{f}u,v\right]^2 \sum_{x,y}\left[t(x-u,y-v) - \bar{t}\right]^2\right\}^{0.5}} \quad (3.3)$$

where $\gamma(u,v)$ is a matrix that contains the correlation coefficients, which can range in value from -1.0 to 1.0. f is the 21-by-21 region around the base control point and the sum is over x,y under the window containing the 11-by-11 template, t, positioned at (u,v). $\bar{t}$ is the mean of the template and $\overline{f}u,v$ is the mean of $f(x,y)$ in the region under the template.

1.3. Find the absolute peak of the cross-correlation matrix and

1.4. Use the position of the peak to adjust the coordinates of the input control point.

Table 3.2 below presents comparison of control point values on the input image before and after the tuning process.

Table 3.2 The result of tuning process

Template	Control Point Locations in X and Y Axes					
	Before		After		Difference	
	Xin	Yin	Xadj	Yadj	X	Y
1	524	148	523.9	149.9	0.1	-1.9
2	356	317	354.1	314.7	1.9	2.3
3	277	287	277	287	0	0

As seen in Table 3.2, the control point extracted from the template 1 has moved to a down direction around 2 pixels (i.e., -1.9), the control point found from the template 2 has moved to an up direction around 2 pixels, 2.3 in Y axis, and the right direction around 2 pixels, 1.9 in X axis and the control point on the template 3 does not move.

Next, control points of the reference image and the new adjustment control points provided by the tuning process were used to estimate parameters of similarity transformation such as scaling, rotating and translating in order to register the IdII-7 into the same coordinate system with the reference image. Fig. 3.9 shows the registered IdII-7.

Next, a subtracted image was produced by subtracting the registered IdII-7 and the reference image. Fig. 3.10, below, shows the subtracted image. As can be seen in Fig. 3.10, there are two areas: a black area and a gray area. The black area is an overlapping area, referred to as region of interest, as a result of simply differencing between the reference image and the registered IdII-7. The gray area is produced by subtracting the reference image and zero values in the registered IdII-7. The black area also contains significant changes. Hence, searching of significant changes was only concentrated on the region of interest.

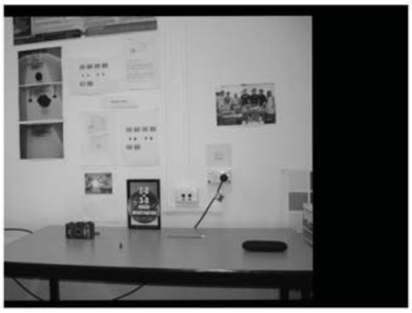


Fig. 3.9 The registered IdII-7. As seen in Fig. 3.9, occluded regions in the right side of the image has been automatically removed as a result of image registration

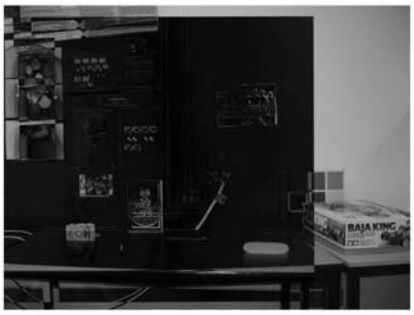


Fig. 3.10 The subtracted image produced after subtracting the reference image and the IdII-7

Next, the region of interest was automatically separated from the subtracted image by using information of the registered IdII-7. The registered IdII-7 was firstly changed into a binary image by using the rule below.

$$RB(x,y)=\begin{cases} 1 & if\ P(x,y) \rangle\ 0 \\ 0 & if\ P(x,y) = 0 \end{cases} \tag{3.4}$$

where RB(x, y) is a binary image of the registered IdII-7 and P(x,y) represents a pixel value for coordinates (x, y). As a result of the image thresholding, the binary image contained white and black areas. The white area was referred to as the bounding box. The bounding box parameters such as the coordinate of the upper left corner, width and height were secondly extracted from the RB(x, y). Based on these bounding box parameters information, the region of interest was finally cropped from the subtracted image.

By applying a global threshold value towards the region of interest, a preliminary change mask was generated. Fig. 3.11 shows the preliminary changed mask.

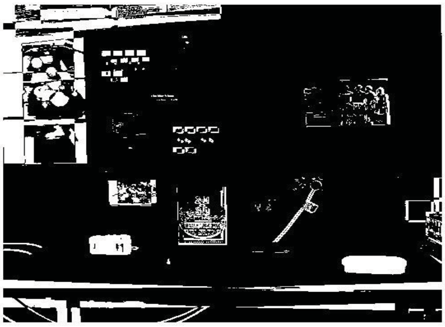


Fig. 3.11 The preliminary changed mask

As can be seen in Fig. 3.11, unimportant and significant changes apparently appear in the preliminary changed mask. Misalignment of the registered IdII-7 seems to be unavoidable and a major cause of unimportant changes appearing in it besides noise and shadow. Two images of the same scene captured by a mobile-camera from two different positions cause a large disparity between pixels of a same object in both images. Registering such a pair of depth plane images by using similarity transformation parameters estimated from the three pairs of correspondence points might not accurately handle pixels which have such large disparity. However, any new object that entered to the scene apparently appeared in the preliminary changed mask.

To separate significant changes while rejecting unimportant changes from the preliminary change mask, unimportant changes removal was then employed. First, any object connected to the preliminary changed mask border was suppressed. Referring to equation (3.5), morphological reconstruction based on dilation is used to suppress light objects connected to the image border. Morphological reconstruction has three unique properties: (1) Processing is based on two images, a marker image and a mask image. In this process, the preliminary changed mask is used as a mask image. Output of morphological reconstruction is called as the marker image, f_m . (2)

Processing repeats until stability; i.e., the image no longer changes. (3) Processing is based on the concept of connectivity and 8-connected neighbourhood is used in this process. Fig. 3.12 shows the result of suppression of light objects connected to the image border.

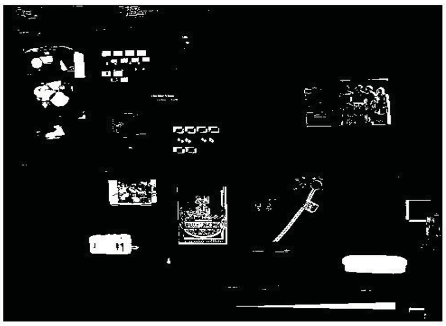


Fig. 3.12 The result of suppression of light objects connected to the image border

Second, erosion was applied towards the image in Fig. 3.12. Horizontal and vertical lines were used as structuring elements in order to reduce, horizontally and vertically, objects in the image. Fig. 3.13 shows the result of this erosion.

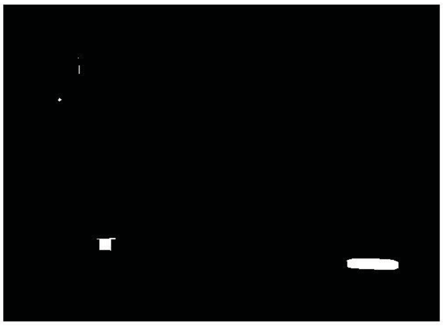


Fig. 3.13 The result of vertical and horizontal erosion

Finally, objects with less than 100 pixels were removed from the image in Fig. 3.13. The result of this process is the final changed mask. Fig. 3.14 shows the result of removing these small objects.

To find the location of these significant changes in the registered IdII-7, centroids, the center of mass of the region, of final objects in the final changed mask were finally extracted. Based on information of these centroids, circles can be drawn on the registered IdII-7 in order to highlight locations of significant changes on the registered IdII-7. Fig. 3.15 shows locations of significant changes on the registered IdII-7.

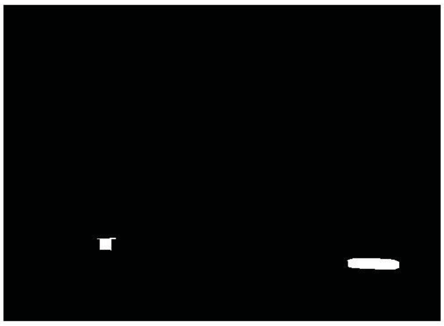


Fig. 3.14 The final changed mask produced after removing small objects less than 100 pixels

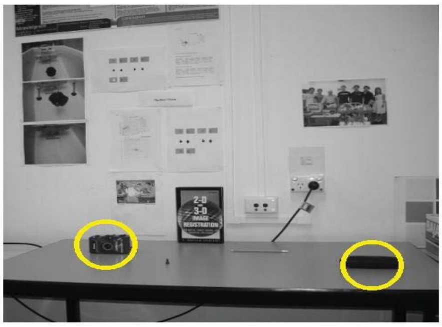
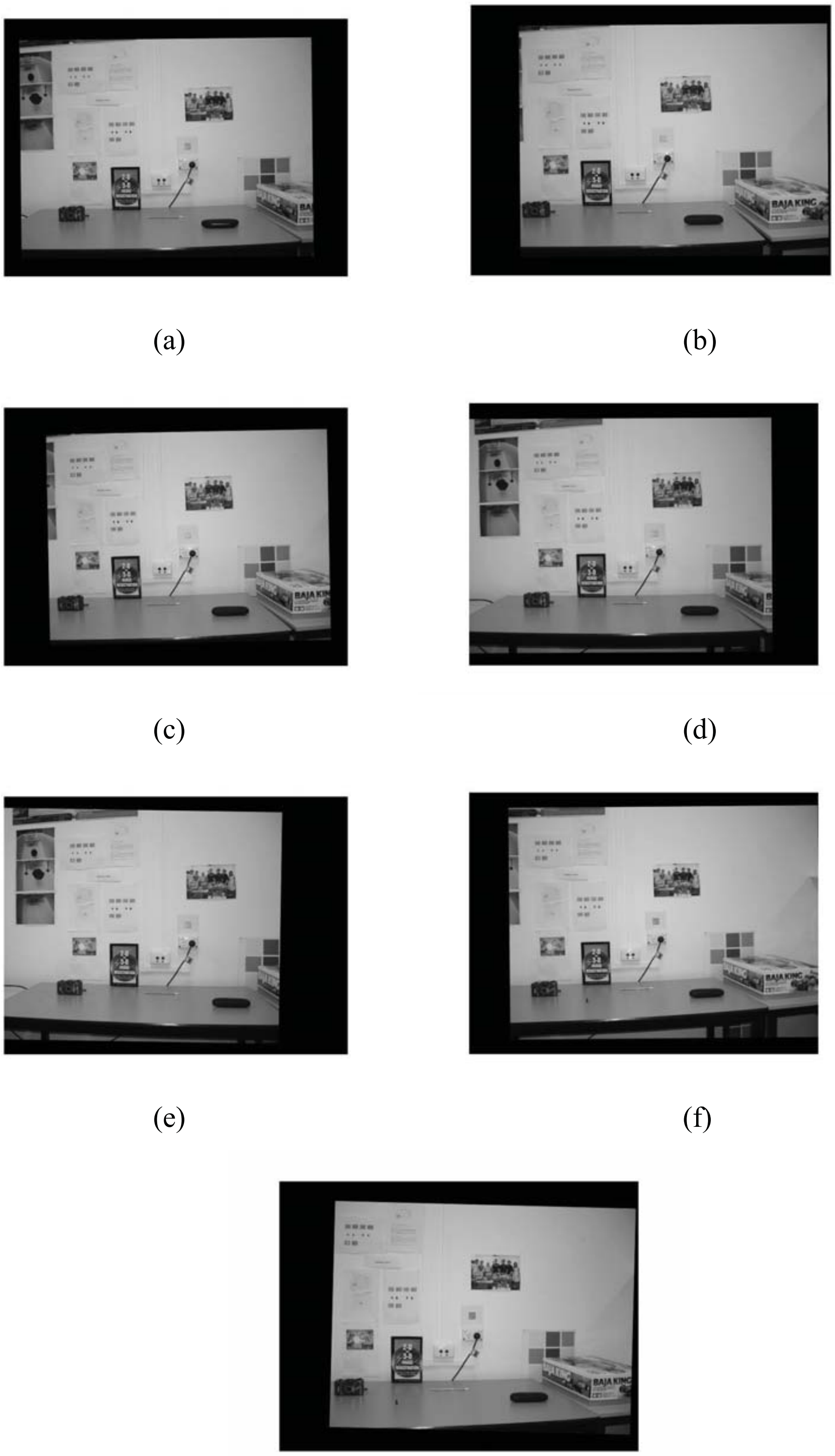


Fig. 3.15 Locations of significant changes on the registered IdII-7

As can be seen in Fig. 3.15, two of the four expected significant changes have been detected correctly by the indoor change detection method. Although the indoor change detection method has successfully detected the significant changes, it has a limitation. The limitation of this method is that the method fails to detect small objects which have less than or equal to 100 pixels or a diameter less than or equal to 1 cm as a consequence of using erosion in the sub-stage of unimportant changes removal.

Furthermore, other input images, IdII-1, 2, 3, 4, 5, 6 and 8, were fed in the indoor change detection method. Figs. 3.16 (a) – (g) depict results of performing automatic image registration towards other input images. As seen in Figs. 3.16 (a) – (g), original input images have been transformed by the similarity transformation in different translation, rotation and scaling based on information of correspondence points extracted from the input images and the reference image.



(a) (b)

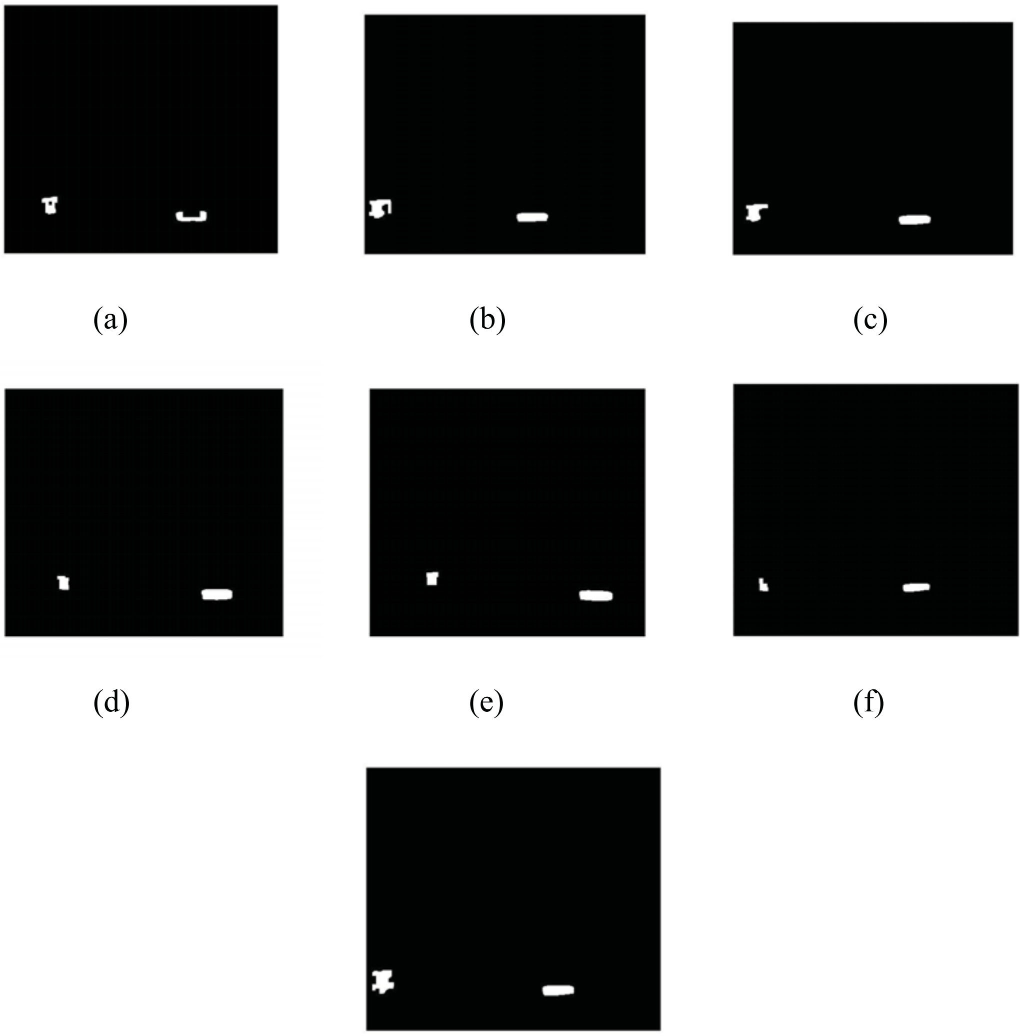
(c) (d)

(e) (f)

(g)

Fig. 3.16 Results of applying automated image registration towards the IdII-1 (a), IdII-2 (b), IdII-3 (c), IdII-4 (d), IdII-5 (e), IdII-6 (f) and IdII-8(g)

Figs. 3.17 (a) – (g) depict changed masks after applying the indoor change detection method towards other input images. As seen in Figs. 3.17 (a) – (g), two big significant changes (i.e., a glass case and conventional camera) apparently appear in all changed masks and two small significant changes (i.e., a pencil and bolt) miss in all changed masks. Moreover, contours of both significant changes do not represent their original contours.



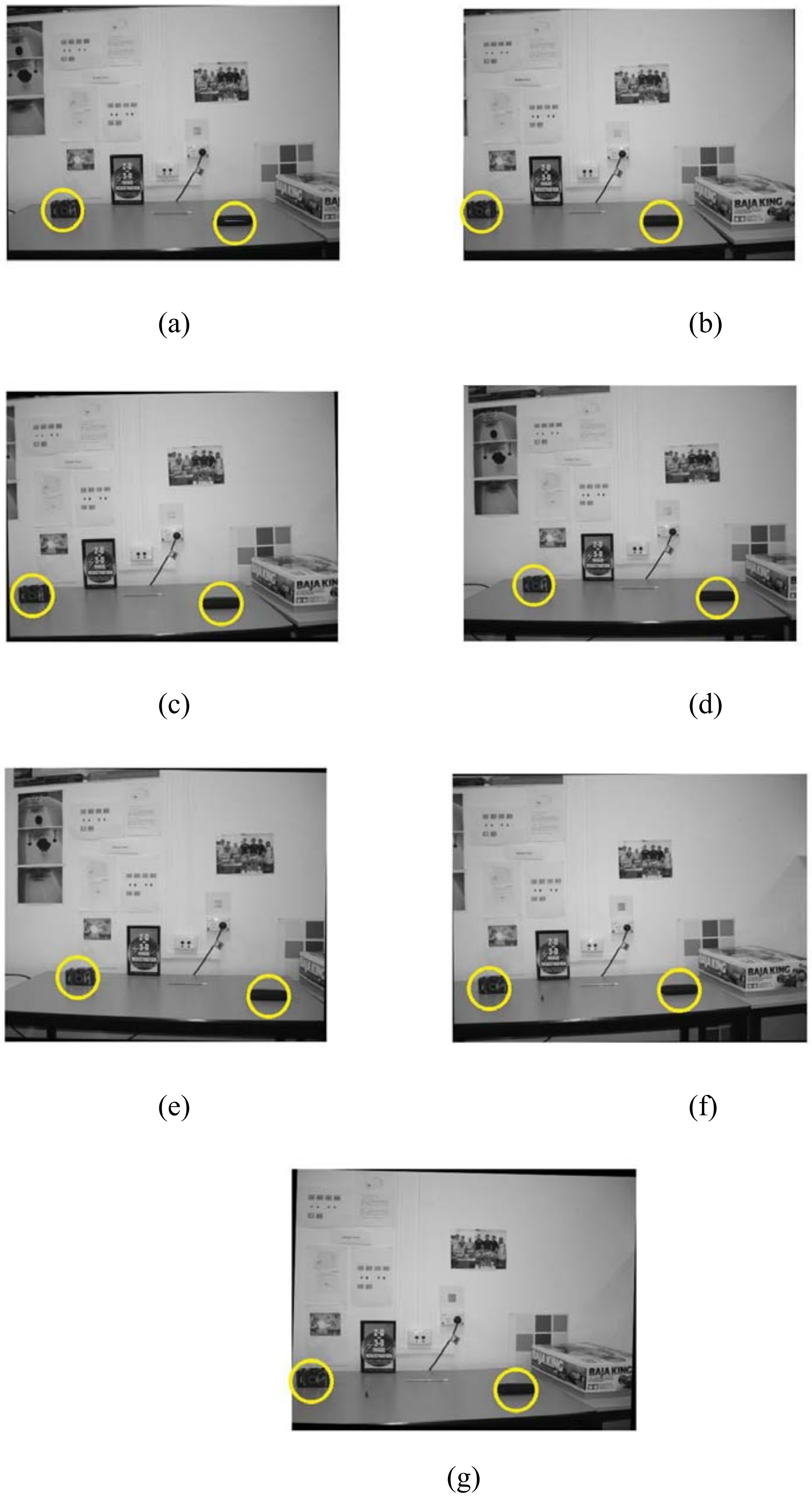
(a) (b) (c)

(d) (e) (f)

(g)

Fig. 3.17 Changed masks of the IdII-1 (a), IdII-2 (b), IdII-3 (c), IdII-4 (d), IdII-5 (e), IdII-6 (f) and IdII-8(g) generated by the indoor change detection method

Fig. 3.18 (a) – (g) depict results of detecting locations of significant changes on registered input images based on information of changed masks.



(a) (b)

(c) (d)

(e) (f)

(g)

Fig. 3.18 Locations of significant changes detected by the indoor change detection method from the IdII-1 (a), IdII-2 (b), IdII-3 (c), IdII-4 (d), IdII-5 (e), IdII-6 (f) and IdII-8 (g)

3.4.1 Subjective Evaluation

Since this indoor experiment is an initial study on change detection methods for multiple images of a same scene captured by a mobile camera from slightly different viewing positions and angles, the objective evaluation of the segmentation quality like ground truth based measures might not be suitable in this initial study. Complexity of these kinds of multiple images, caused by parallax and changes in object's sizes as a result of camera motion, and artefact lights such shadows and overcontrast regions as a result of illumination variation, is the main reason of why the objective evaluation is not applicable. Thus, the subjective evaluation by human observers is used in this initial study.

Table 3.3 summarizes change detection results produced by the indoor change detection method towards all input images.

Table 3.3 Summarization of change detection results

Input Images	**Expected TP**	**TP**	**FN**	**FP**
IdII-1	4	2	2	-
IdII-2	4	2	2	-
IdII-3	4	2	2	-
IdII-4	4	2	2	-
IdII-5	4	2	2	-
IdII-6	4	2	2	-
IdII-7	4	2	2	-
IdII-8	4	2	2	-
Total	32	16	16	0

where Expected TP is the number of correctly significant changes, TP stands for true positive, correctly detected as significant changes, FN is false negative (miss), falsely detected as significant changes and FP is false positive (false alarm), falsely marked as significant changes.

The true positive rate (TPR) and false negative rate (FNR) are determined by referring to equations 3.5 and 3.6, below.

$$TPR = TP/(Expected\,TP) \quad (3.5)$$

$$FNR = FN/(Expected\,TP) \quad (3.6)$$

The TPR and FNR of the indoor change detection method are 50 % and 50 %. The method misses two small objects which have size less than 100 pixels (i.e., objects with less than 1 cm in diameter when input images were captured from 230 cm distance from the digital camera). In the other hand, the method does not detect any unimportant changes at all.

3.5 Concluding Remarks

Conclusions that can be drawn from the initial study are that:

1. Parallax can be reduced by registering a current input image into the same coordinate system with the reference image. However, prior to register both input and reference images, correspondence points have to be extracted automatically from both input and reference images. Extracting control points from both input and reference images is an issue since sizes and pixel values of same objects in both input and reference images are significantly changes because of camera movement and illumination variation.
2. Since images used in this preliminary study are depth plane images (i.e., images were captured in such a way that the position of a digital camera was perpendicular towards the scene), registering two depth plane images will only bring regions which are at and/or near control points into the same coordinate system (zero and/or smaller disparity values) and regions which are far away from control points have larger disparity values.
3. Applying a morphological operation like erosion will remove unimportant changes from a changed mask. However, the erosion may also eliminate significant changes in the changed mask at the same time.

Conclusions drawn from the preliminary study will be used as an additional knowledge in developing an automated outdoor image change method in the next experiment. Moreover, the indoor change detection method has been presented in the 2007 Australasian Conference on Robotics and Automation (ACRA 2007) (Tanjung and Lu, 2007).

CHAPTER 4

AUTOMATIC CONTROL POINTS EXTRACTION

Chapter 4 describes how to extract automatically control points (CPs) from both reference and input images. This chapter also explains how reference and input images used in this outdoor experiment were captured by a mobile camera and presents experimental results of automatic CPs extraction.

4.1 Background

Prior to register automatically two images of the same scene captured by a static or mobile camera, CPs must be extracted automatically from the two images. CPs are pixels whose coordinates are identified in both images. Methods for extracting CPs in two images of the same scene are classified into two different algorithms: area-based and feature-based algorithms (Orduyilmaz, 2006). Area-based algorithms measure statically the similarity of a small window of points extracted from the first image also known as the reference image with a small window of points extracted from the second image referred to as the input image (Fonseca and Manjunath, 1996). Feature-based algorithms match a set of features extracted from the reference image and the input image. The features (e.g., edges, regions, lines, line endings, line intersections or region centroid) must be invariant towards rotation, scaling and varying illumination (Sester, Hild, and Fritsch, 1998; Roux, 1996).

Extracting automatically CPs in multiple outdoor images of the same scene containing fence wires acquired by a mobile camera is a difficult task. Area-based algorithms may fail to extract CPs from these kinds of multiple outdoor images because size, position and intensity values of same objects change significantly in the multiple outdoor images. Features like edges, lines and corners are also very difficult to be detected in these kinds of multiple outdoor images because of the complexity of background clutter of the scene, illumination variation during the day and camera movement. Therefore, local image features referred to as SIFT keys proposed by Lowe (2004) are utilized in this research. SIFT stands for Scale Invariant Feature Transform. SIFT keys are invariant to image scaling, translation, rotation and partially invariant to illumination changes and affine or 3D projection. A completed explanation of the SIFT operator is presented in appendix 1.

4.2 Algorithm Overview

In extracting automatically CPs from these kinds of multiple outdoor image of the same scene, a template-based matching approach is developed in this research. Two similar human-made templates (e.g., building pictures) that contain lots of SIFT keys was attached in front of posts of fence wires. The difficulty of extracting control points from natural templates in an outdoor scene because of background clutter, fence wires (i.e., fence wires can separate a big object in the scene into several small parts) and artefact lights produced by illumination changes is the main reason of attaching these artificial templates. Furthermore, a human-made template (template_RI) is manually cropped from the reference image in advance. The followings are steps established in extracting CPs from reference and input images.

For every reference or input image,

1. Crop every original reference or input image to a specified rectangle, referred to as an area of interest of reference image (AIO1_RI) or an area of interest of input image (AIO1_II). The rectangle is a four-element horizontal vector with the form [xmin ymin width height)]; these values are specified in spatial coordinates and obtained by using the following equations. This step will remove automatically the top of each original reference or input image, which contains mostly trees and the sky. Intruders and any suspicious objects put by intruders within the perimeters of protected areas are commonly being on or close to the ground; hence,

searching of significant changes is only focus from the ground to 1.5 meters height.

$$x\min = 1 \quad (4.1)$$

$$y\min = h/2 \quad (4.2)$$

$$y_2 \min = y\min/2 \quad (4.3)$$

$$width = w \quad (4.4)$$

$$height = y\min + y_2 \min \quad (4.5)$$

in which h and w are the height and width of the reference image in pixels. The size of images used in this study was an 8Mb file (3264 by 2448 pixels). Hence, h is 2448 pixels and w is 3264 pixels.

2. Divide the AIO1_RI or the AIO1_II into four separated regions: I, II, III and IV regions (see Fig. 4.1 below).
3. For extracting a control point (CP) from the left post of fence wires of the AIO1_RI,
   - 3.1 Read the AIO1_RI and the template_RI.
   - 3.2 Perform the SIFT operator towards the template_RI and only region I of the AIO1_RI.
   - 3.3 Extract only matched keypoint locations in the region I of the AIO1_RI.
   - 3.4 Calculate an average value from the extracting matched keypoint locations and use the average value as the first control point.
4. For extracting a CP from the right post of fence wires of the AIO1_RI,
   - 4.1 Read the AIO1_RI and the template_RI.
   - 4.2 Perform the SIFT operator towards the template_RI and only region IV of the AIO1_RI.
   - 4.3 Extract only matched keypoint locations in the region IV of the AIO1_RI.
   - 4.4 Calculate an average value from the extracting matched keypoint locations and use the average value as the second control point.

5 For extracting control points from the left and right posts of fence wires of the AIO1_II, repeat steps 3 and 4 in which the AIO1_RI is changed with the AIO1_II.

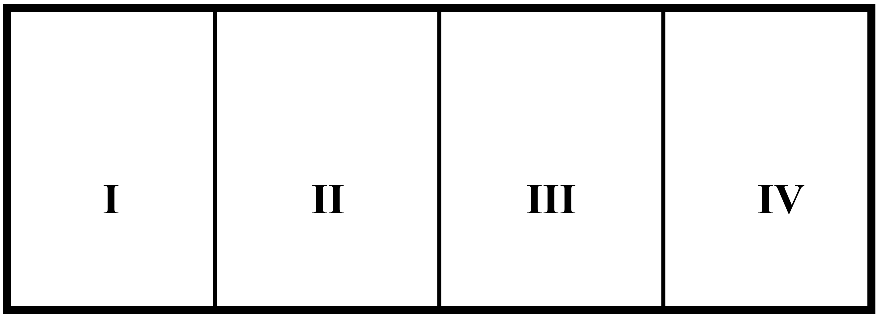


Fig. 4.1 The AIO1_RI and AIO1_II are separated into four regions: I, II, III and IV. Regions I and IV contain the left post and right post of fence wires. Searching of SIFT keys is only performed in regions I and IV in order to reduce errors of SIFT keys matching between SIFT keys extracted in the template_RI and SIFT keys detected in regions I and IV of the AIO1_RI or the AIO1_II

4.3 Reference and Input Images

In order to test the robustness of the automatic CPs extraction algorithm, 13 outdoor images of the same scene containing fence wires including a reference image and 12 input images acquired by a mobile camera from slightly different positions, angles and at different times were utilized. As mentioned in Section 1.2, the movement of the mobile camera is restricted in three degrees of freedom (3DOF) which are X, Y and $\Delta\theta_z$ axes. The camera movement in X and Y axes is limited in ± 20 cm and the rotation of the mobile camera in Z axis is limited in ± 15 degrees.

Fig. 4.2, below, depicts top views of the mobile camera positions when capturing the 13 outdoor images of the same scene containing fence wires. The red colour symbol (🟥) in Fig. 4.2 represents a position in which a reference image was captured and the ■ symbol represents the mobile camera. The position of capturing the reference image is referred to as the origin (0) position (0, 0, 0). In addition, the reference image was captured on 25th of December 2008 at 12.21 PM in the spring time. Fig. 4.3 depicts the reference image used in this research.

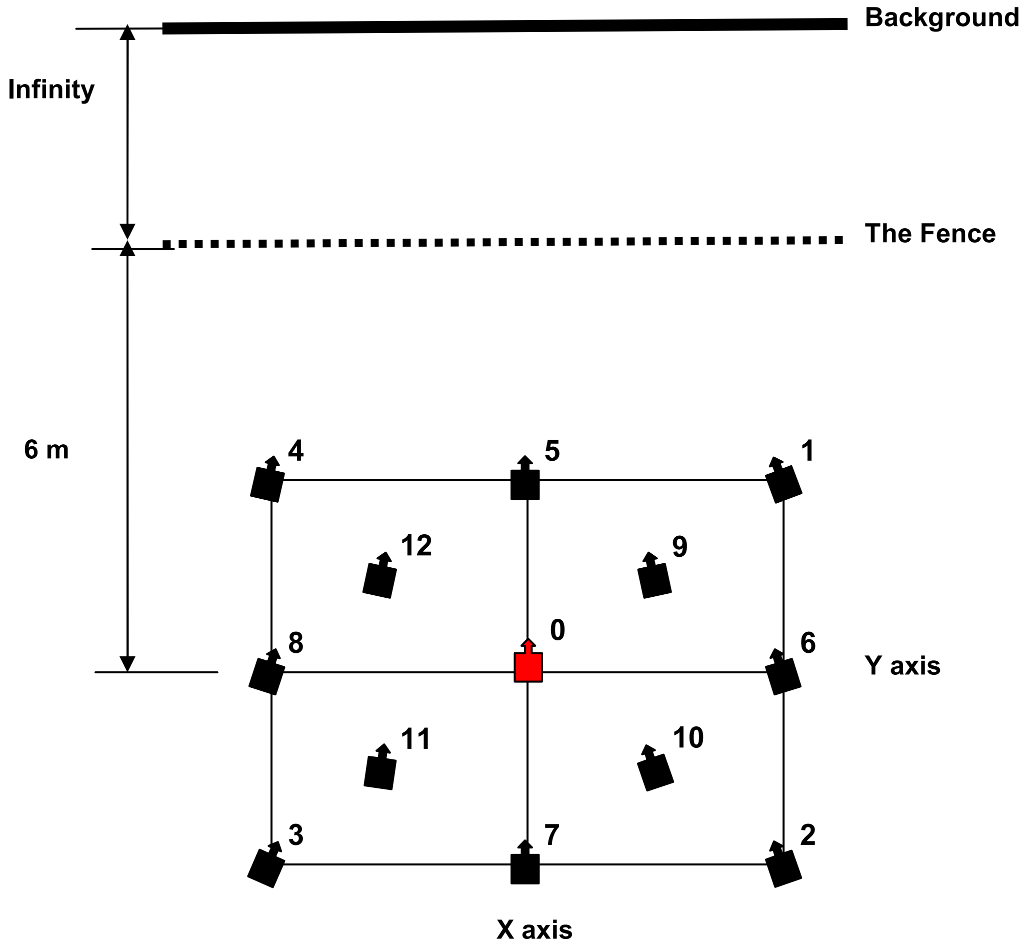


Fig. 4.2 Top views of the mobile camera positions when capturing the 13 outdoor images of the same scene containing fence wires

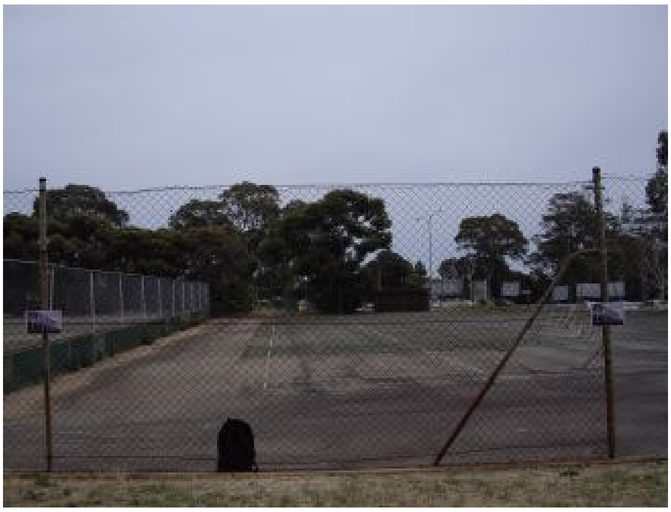
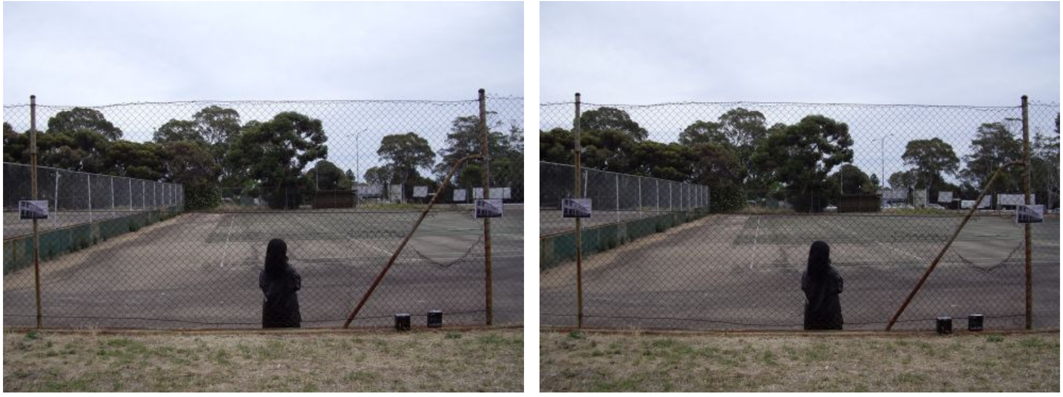
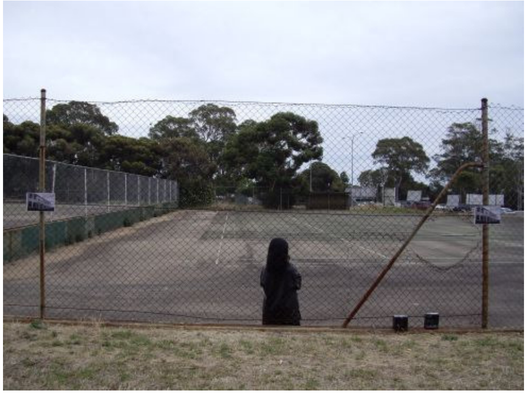


Fig. 4.3 The reference image used in this outdoor experiment

After capturing the reference image, three new objects (i.e., two small boxes in front of the wire fence and someone who represents an intruder behind the wire fence) were added to the scene and an object (i.e., a school bag behind the wire fence that denotes a disappearing old object) was removed from the scene. In addition, two breaches in the integrity of the wire fence (i.e., a large breach in the middle right of and a tiny breach in the bottom right of the wire fence) were added to the scene. The digital camera was then shifted manually 20 cm forward (zoom in) and 20 cm to the right. It was then rotated 15 degrees in an anticlockwise direction. This current camera position, (+20, +20, -15), was referred to as the 1st input camera position, from which the reference image was captured, the input image 1 (see Fig. 4.2, above). Position 1 in Fig. 4.2, above, shows a top view of the 1st input camera position. The input image 1 (II-1) was captured on 25th of December 2008 at 02.51 PM. To capture the input image 5 and the input image 9 (II-5 and II-9), the mobile camera was then shifted manually to the 5th input camera position (0, +20, 0) and 9th input camera position (+10, +10, -10), from which the reference image was captured. Position 5 and position 9 in Fig. 4.2, above, show top views of the 5th and 9th input camera positions. Both the II-5 and the II-9 were captured at slightly different times with the II-1 (i.e., several seconds after capturing the II-1). The reason of capturing the II-1, II-5 and II-9 at slightly different times is only to simulate that the mobile camera could probably come into these positions in the real time application. Figs. 4.4 (a), 4.4 (b) and 4.4 (c) depict the II-1, II-5 and II-9.



(a) (b)

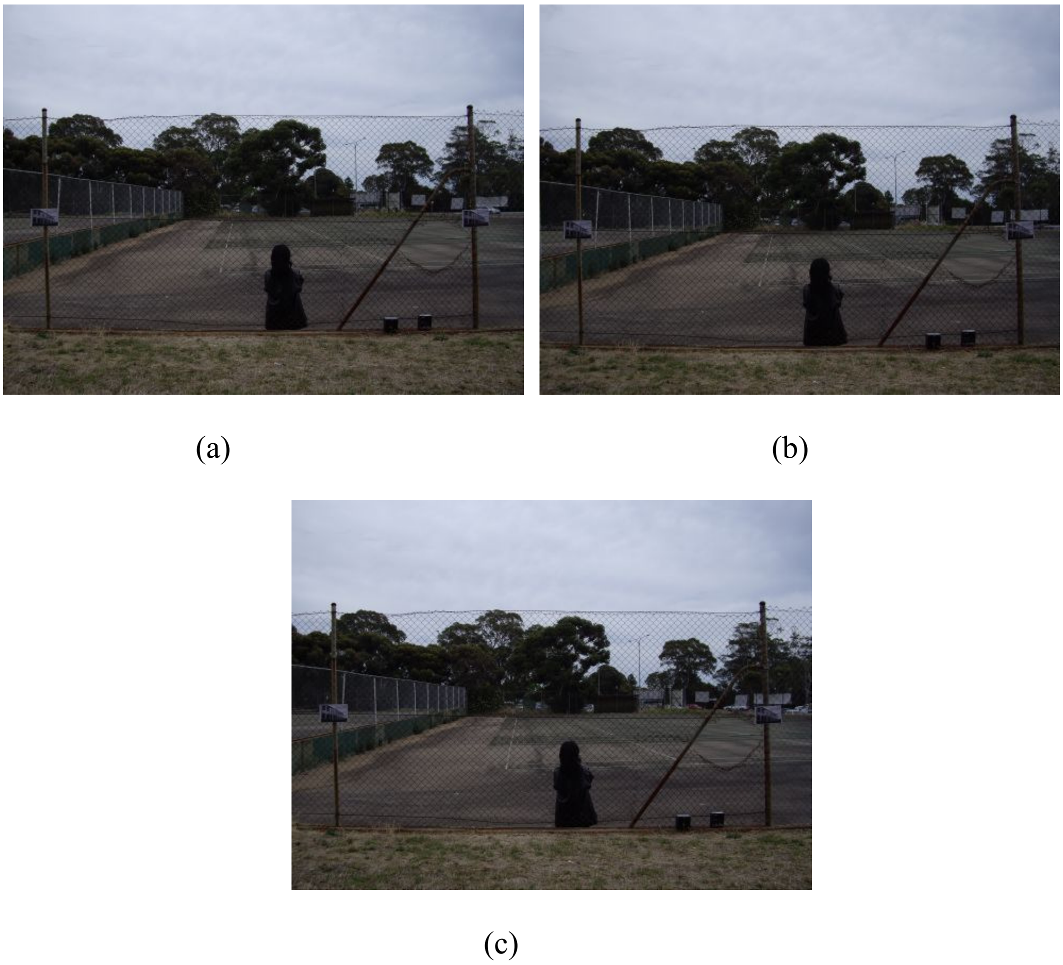


(c)

Fig. 4.4 The II-1 (a), II-5 (b) and II-9 (c) were captured by a mobile camera at slightly different positions, angles and at slightly different times. These input camera positions simulate that the mobile camera could be in one of these positions in the real time application

After capturing the II-1, II-5 and II-9, the mobile-camera was put back to the origin position. From the origin position, the digital camera was manually moved again 20 cm backward (zoom out) and 20 cm to the right. It was then turned 15 degrees in an anticlockwise direction. This current camera position, (+20, -20, -15), was referred to as the 2nd input camera position. From this position, the II-2 was then captured. Position 2 in Fig. 4.2, above, depicts a top view of the 2nd input camera position. The II-2 (input image 2) was captured on 25th of December 2008 at 04.57 PM. To capture the input image 6 and the input image 10 (II-6 and II-10), the mobile camera was then shifted manually to the 6th input camera position (+20, 0, -10) and 10th input camera position (+10, -10, -10), from which the reference image was captured. Position 6 and position 10 in Fig. 4.2, above, show top views of the 6th and 10th input camera positions. Both the II-6 and the II-10 were captured at slightly different times with the II-2 (i.e., one minute after capturing the II-2). The reason of capturing the II-2, II-6 and II-10 at slightly different times is only to simulate that the mobile camera could probably come into these positions in the real time application. Figs. 4.5 (a), 4.5 (b) and 4.5 (c) depict the II-2, II-6 and II-10, respectively.

The II-3 was captured at a (-20, -20, +15) position from the origin position on 25th of December 2008 at 7.23 PM. The II-7 and II-11 were captured at slightly different times from the II-3 (i.e., one minute after capturing the II-3) at (0, -20, 0) and (-10, -10, +10) positions. Positions 3, 7 and 11 in Fig. 4.2, above, depict top views of the 3rd, 7th and 11th input camera positions. Figs. 4.6 (a), 4.6 (b) and 4.6 (c) below show the II-3, II-7 and II-11.



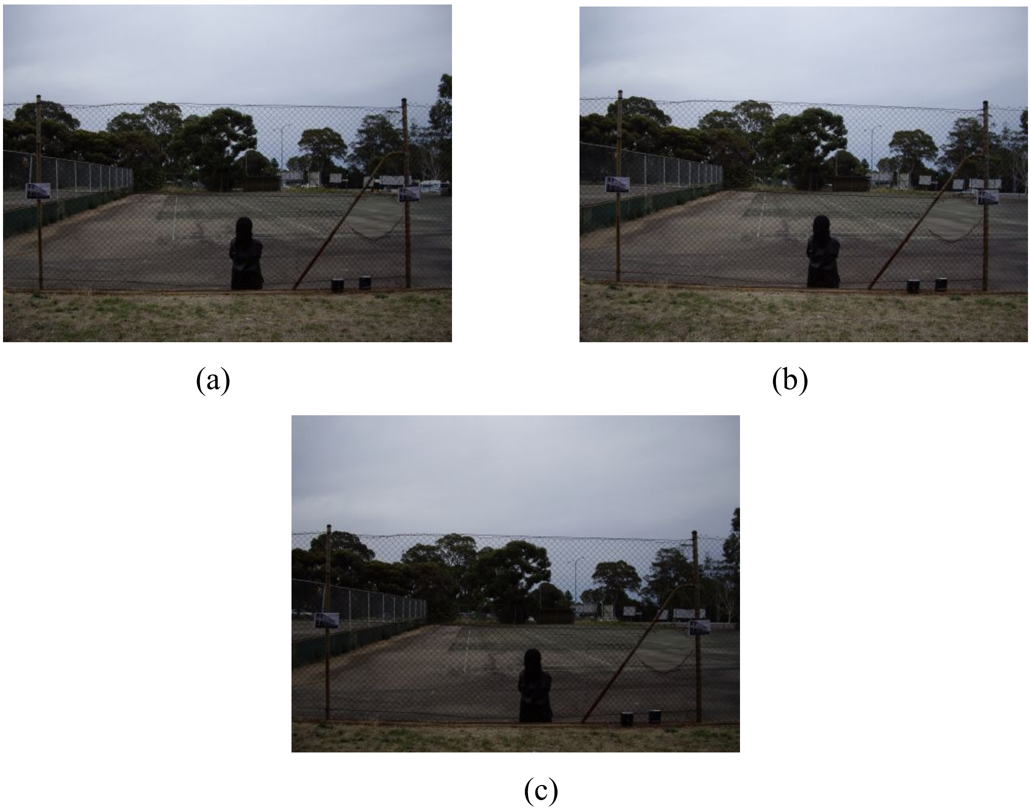
(a)

(b)

(c)

Fig. 4.5 The II-2 (a), II-6 (b) and II-10 (c) were captured by a mobile camera at slightly different positions, angles and at slightly different times. These input camera positions simulate that the mobile camera could be in one of these positions in the real time application

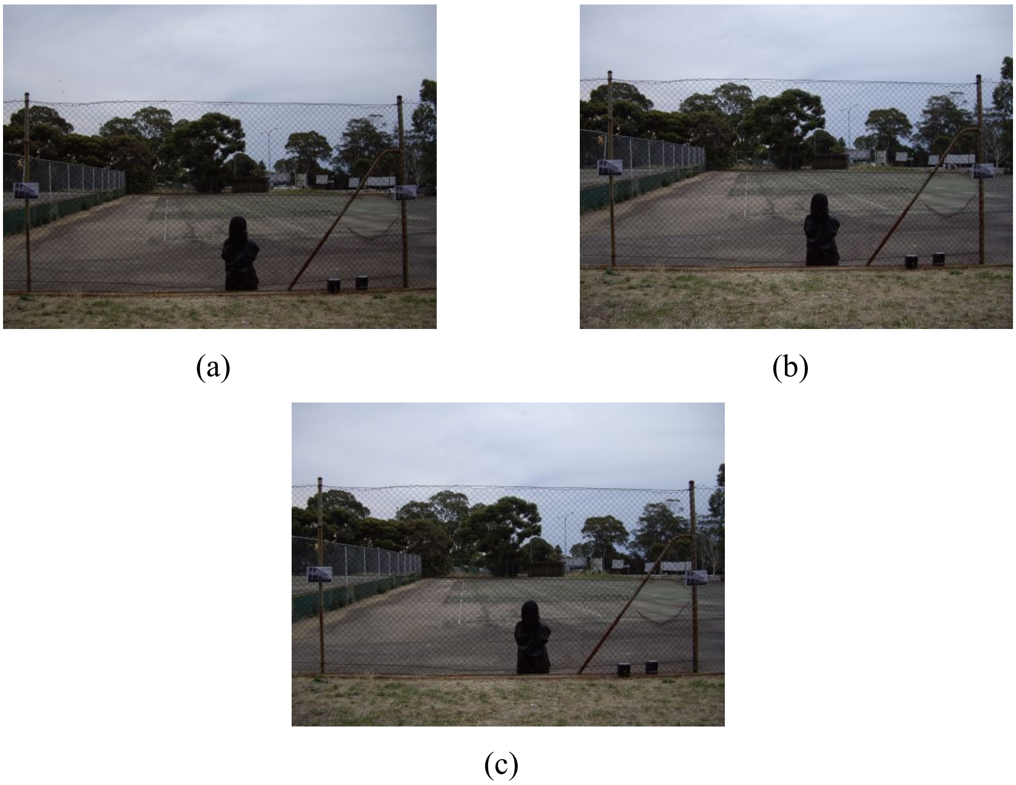
The II-4 was acquired at a (-20, +20, +25) position from the origin position on 25th of December 2008 at 7.41 PM. Several seconds later, the II-8 and II-12 were captured at (-20, 0, +10) and (-10, +10, +10) positions from the origin position. Positions 4, 8 and 12 in Fig. 4.2, above, depict top views of the 4th, 8th and 12th input camera positions. Figs 4.7 (a), 4.7 (b) and 4.7 (c) below show the II-4, II-8 and II-12.



(a) (b)

(c)

Fig. 4.6 The II-3 (a), II-7 (b) and II-11 (c)



(a) (b)

(c)

Fig. 4.7 The II-4 (a), II-8 (b) and II-12 (c) were acquired from slightly different positions, angles and at slightly different times

Table 4.1, below, summarizes positions and times of capturing all images including the reference image and input images used in this outdoor experiment. All images were captured on 25th of December 2008.

Table 4.1 Summarization of positions and times when capturing the reference image and input images used in this research

Images	**Positions** **(X, Y, $\Delta\theta_z$)**	**Times** **(hr:mn:sc) PM**	**Positions** **in Fig. 2**
Reference Image	(0, 0, 0)	12:21:00	0
Input Image 1 (II-1)	(+20, +20, -15)	02:51:00	1
Input Image 5 (II-5)	(0, +20, 0)	02:51:25	5
Input Image 9 (II-9)	(+10, +10, -10)	02:51:55	9
Input Image 2 (II-2)	(+20, -20, -15)	04:57:00	2
Input Image 6 (II-6)	(+20, 0, -10)	04:58:00	6
Input Image 10 (II-10)	(+10, -10, -10)	04:58:10	10
Input Image 3 (II-3)	(-20, -20, +15)	07:23:00	3
Input Image 7 (II-7)	(0, -20, 0)	07:24:00	7
Input Image 11 (II-11)	(-10, -10, +10)	07:24:15	11
Input Image 4 (II-4)	(-20, +20, +15)	07:41:00	4
Input Image 8 (II-8)	(-20, 0, +10)	07:41:25	8
Input Image 12 (II-12)	(-10, +10, +10)	07:41:55	12

4.4 Experimental Results and Discussion

To solve the problem of extracting CPs in multiple images of the same outdoor scene, a template-based matching approach by using the SIFT operator is proposed in this research. A human-made template (template_RI) was manually cropped from the reference image in advance. Fig. 4.8, below, depicts the template_RI.

NOTE:
This figure is included on page 55
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 4.8 A human-made template is picture of a building

The reason of choosing this school building picture is that corners of the building provide SIFT keys that are robust towards illumination variation, rotation, scaling and affine or 3D projection. From this template, the SIFT operator can extract 252 SIFT keys (keypoints).

The following is an explanation of extracting two CPs from the reference image. According to the first step of extracting automatically CPs algorithm in Section 4.2 above, the reference image is cropped into a specific rectangle in order to remove the sky and trees in the background. The cropped reference image is also called as the first area of interest of the reference image (AOI1_RI). Fig. 4.9, below, depicts the AOI1_RI.

Next, the AOI1_RI is automatically divided into four regions: I, II, III and IV. Figs. 4.10 (a) and 4.10 (b) depict only region I and region IV which contain left and right posts of the wire fence. The SIFT operator is firstly applied to the template_RI and region I of the AOI1_RI in searching matched keypoints and it extracts 126 matched keypoints. Fig. 4.11, below, depicts matched keypoints extracted from the template_RI and region I of the AOI1_RI. An average value of matched keypoint locations only in region I of the AOI1_RI is then calculated and it is used as a first CP.

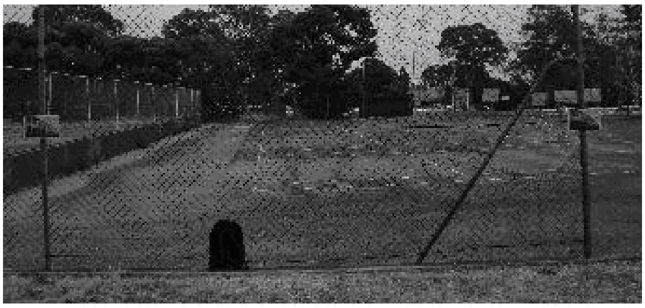
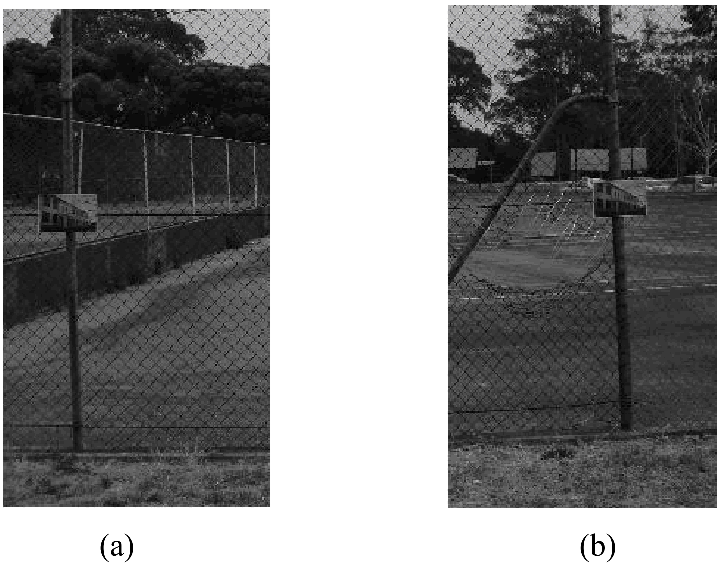


Fig. 4.9 The AOI1_RI cropped automatically from the reference image



(a) (b)

Fig. 4.10 Regions I (a) and IV (b) cropped automatically from the reference image. Both regions contain left and right posts of the wire fence

The SIFT operator is performed again to the template_RI and region IV of the AOI1_RI in searching other matched keypoints. It extracts 23 matched keypoints. Fig. 4.12, below, depicts matched keypoints detected from the template_RI and region IV of the AOI1_RI. The second average value of matched keypoint locations only in region IV of the AOI1_RI is then calculated and this second average values is used as a second CP detected from the right post of the wire fence. Fig. 4.13, below, depicts both CPs extracted from left and right posts of the wire fence in the AOI1_RI.

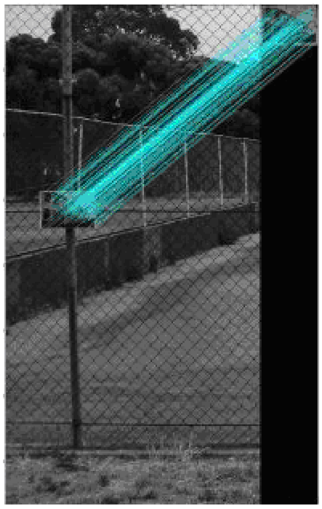


Fig. 4.11 The SIFT operator extracts 126 matched keypoints from the template_RI and region I of the AOI1_RI.

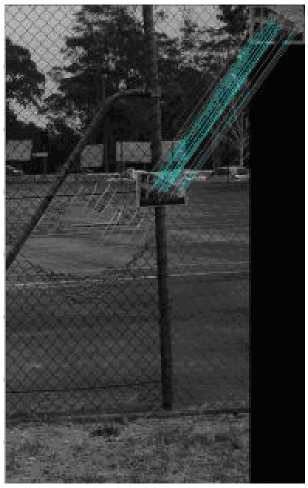


Fig. 4.12 The SIFT operator extracts 23 matched keypoints from the template_RI and region IV of the AOI1_RI

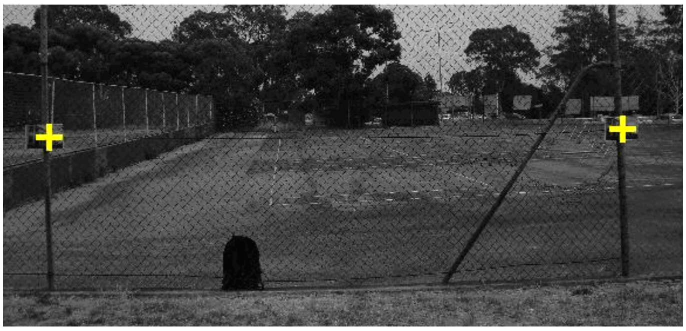
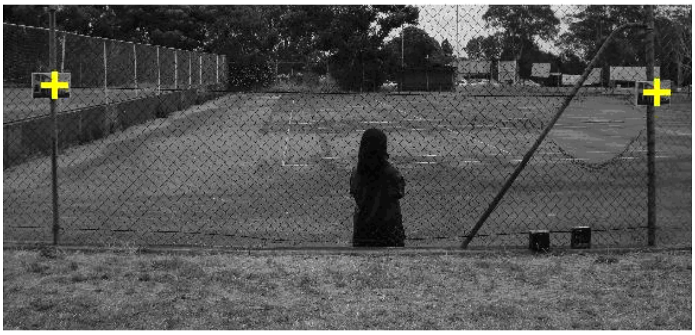


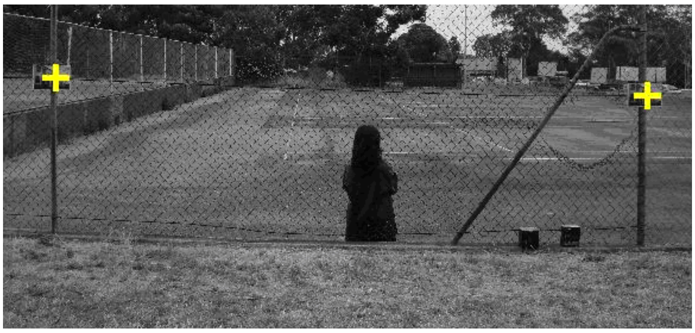
Fig. 4.13 Two CPs are automatically extracted from left and right posts of the wire fence with a template-based matching approach by using the SIFT operator in the AOI1_RI

The same algorithm is applied to all input images. The following Figs. depict CPs extracted automatically from other input images.

(a)

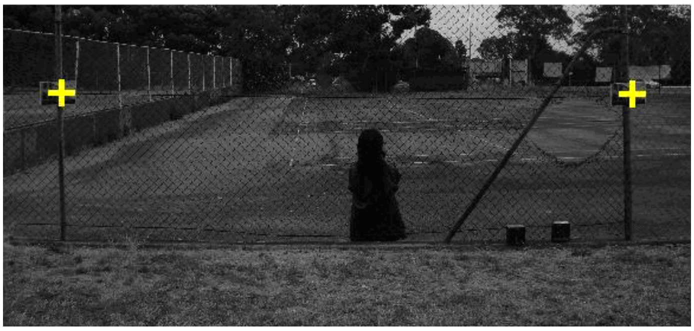


(b)



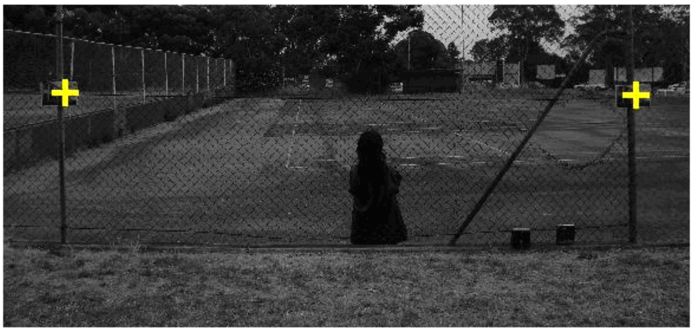
(c)

Fig. 4.14 CPs extracted automatically from the AOI1_II-1 (a), AOI1_II-5 (b) and AOI1_II-9 (c)



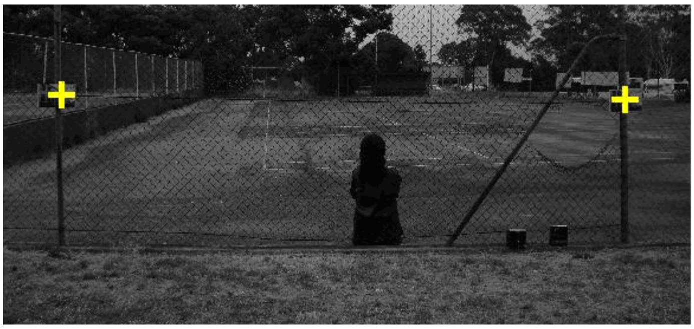
(a)

(b)

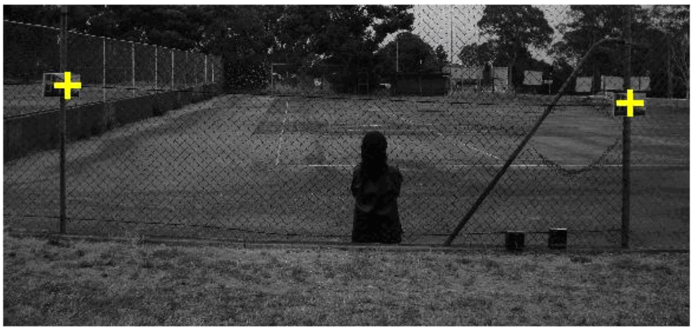


(c)

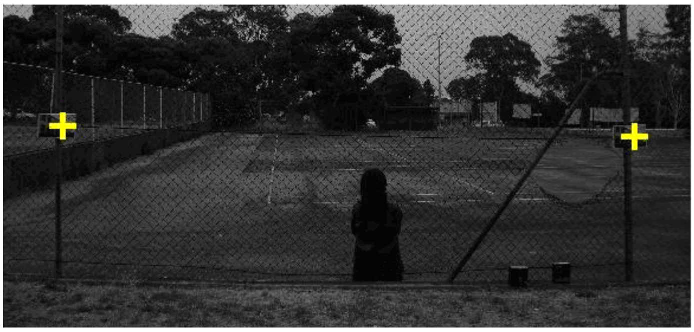
Fig. 4.15 CPs detected from the AOI1_II-2 (a), AOI1_II-6 (b) and AOI1_II-10 (c)



(a)

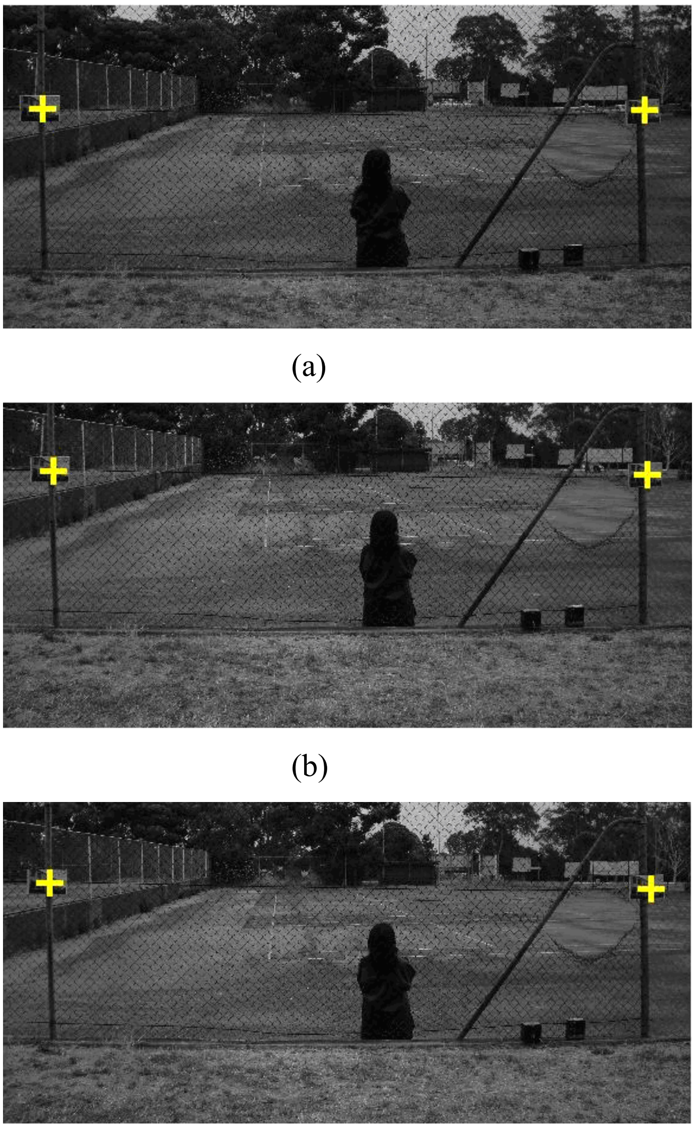


(b)



(c)

Fig. 4.16 CPs detected from the AOI1_II-3 (a), AOI1_II-7 (b) and AOI1_II-11 (c)



(a)

(b)

(c)

Fig. 4.17 CPs detected from the AOI1_II-4 (a), AOI1_II-8 (b) and AOI1_II-12 (c)

Table 4.2, below, summarizes matched keypoints extracted by the SIFT operator from left and right posts of the wire fence in each input image. As can be seen in Table 4.2, matched keypoints provided by the SIFT operator vary in each input image. The varieties occur because of significant changes on illumination, rotation, translation and scaling during capturing input images. However, illumination variations and scaling are two greatest effects in making such varieties of the number of matched keypoints since both illumination variations and scaling have a great consequence on changes of pixel intensity values in each input image.

As seen in Figs. 4.14, 4.15, 4.16 and 4.17, above, CPs extracted from left and right posts of the wire fence are on the human –made template. As long as the CPs are on the template, the CPs are acceptable to be used in registering the reference image and one of input images. When the CPs are outside of the template, the CPs are not suitable to be utilized for image registration since they will make a large error in image transformation and resampling.

Table 4.2 Summarization of matched keypoints extracted from left and right posts of the wire fence in each input image

Input Image (II)	Matched Keypoints	
	Left Post	Right Post
1	44	19
5	41	16
9	44	25
2	31	11
6	28	12
10	31	14
3	18	15
7	20	22
11	11	11
4	23	16
8	20	14
12	17	15

4.5 Concluding Remarks

Extracting automatically CPs for image registration in multiple images of the same outdoor scene containing fence wires captured by a mobile camera from slightly different positions, angles and at different times is a difficult task. Background clutter of the outdoor scene, camera movement, illumination variations and tiny fence wires are factors that make the CPs extracting task becomes very difficult. In addition to the effect of tiny fence wires, tiny fence wires can divide a big object behind the wire fence into several small objects and fence wires are natural objects with less local features.

The template-based matching approach using the SIFT operator has proven that it can be used to extract CPs from these kinds of multiple outdoor images. The success depends on choosing the human-made template. The human-made template must contain potential features that are invariant towards scaling, rotation, affine projection and illumination variations. These kind features can be founded in corners of a big building. This is the reason of choosing a building as template in this research.

Natural templates that can be extracted from the outdoor scene such as posts of the wire fence, trees, buildings or objects on background may not be the best option for this research because these natural templates contain fewer features. Possibility of using these natural templates will be further explored in the future study. In addition, sizes and intensity values of these natural templates change significantly because of camera movement and varying illumination. As a consequence, intensity-based methods may fail to extract CPs in these kinds of multiple outdoor images.