UPPSALA
UNIVERSITET

IT 13 030
Examensarbete 30 hp
Maj 2013

Rotation Invariant Registration of 2D Aerial Images Using Local Phase Correlation

Ping Lu

Institutionen för informationsteknologi
Department of Information Technology

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Rotation Invariant Registration of 2D Aerial Images Using Local Phase Correlation

Ping Lu

Aerial image registration requires a high degree of precision. In order to improve the accuracy of feature-based registration, this project proposes a novel Log-Polar Transform (LPT) based image registration. Instead of using the whole image in the conventional method, feature points are used in this project, which reduces the computational time. For rotation invariance, it is not important how the image patch is rotated. The key is focusing on the feature points. So a circular image patch is used in this project, instead of using square image patches as used in previous methods. Existing techniques for registration with Fast Fourier Transform (FFT) always do FFT first and then Log-Polar Transformation (LPT), but it is not suitable in this project. This project does LPT first and then the FFT.

The proposed process of this project contains four steps. First, feature points are selected in both the reference image and the sensed image with corner detector (Harris or SIFT). Secondly, image patches are created using feature point positions as centers. Each point is a center point of LPT, so circular image patches are cropped choosing a feature point as center. The radius of the circle can be changed. Then the circular images are transformed to Log-Polar coordinates. Next, the LPT images are dealt with using phase correlation. Experimental results demonstrate the reliability and rotation invariance of the proposed method.

Handledare: Anders Hast
Ämnesgranskare: Stefan Seipel
Examinator: Ivan Christoff
IT 13 030
Tryckt av: Reprocentralen ITC

Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I wish to express my deepest gratitude to my supervisors Associate Professor *Anders Hast* for his genuine support, valuable advice and sincere comments which helped me a lot in finishing this project. His excellent scientific expertise, in combination with his friendly and helpful character has made him an outstanding academic teacher.

I am thankful to my examiner Professor *Stenfan Seipel* for his invaluable suggestions. I also want to thank my academic teachers Associate Professor *Cris Luengo*, Associate Professor *Robin Strand*, and Professor *Ingela Nyström* who shared their experiences with me during my master study at Uppsala University.

My deepest gratitude goes to my parents, who always have supported me to pursue my dreams. They showered me with love, kindness and care. Last but not least, I wish to thank my friends and classmates who helped me directly or indirectly to complete this project. Special thanks to *Minpeng Zhu*, *Kim-Anh Tran*, *Maria Carlson*, *Krishna Paudel*, *Jun He*, *Zhu Tang*, *Narges Rajabnezhad* and many more.

Contents

List of Figures

Chapter 1

Introduction

Image Registration is an essential step in most image processing tasks. It is widely used in remote sensing, weather forecasting, medical imagimg and so on. Image Registration is the process of overlaying two or more images that have the same scene taken at different image sensors, different times, and different geometric viewpoints, or share the same visual information such as the same objects. The main goal of image registration is to find the transformation between the reference image and the sensed images. However, it could not design a universal registration system to solve all kinds of tasks depending on the different type of image and various types of degradations.

This chapter introduces the majority of registration methods, describes the characteristic of aerial imaging and discusses the problems involved with aerial imaging. Then it describes the proposed solution and conclude by reviewing the thesis.

1.1 Image Registration Process

Image registration consists of the following four steps.

1. **Feature Extraction**
   This step includes feature construction and feature selection. The feature points (also called *key points*) in the image are found by one of the corner detector methods. Each feature point is unique, which is used for feature matching. The basic idea of corner detector will be described in section 2.2.2.

2. **Feature Matching**
   Feature correspodence between the reference and the sensed images is established in this step. It gives the best matching point pairs between the images using similarity measures.

3. **Homography Estimation**
   Based on the matching points, this step can estimate the parameter of the mapping function. In order to get the parameters, homography estimation only uses the true matching points. If the wrong matching points are used then the images will be incorrectly aligned. The detail of homography estimation will be discussed in section 2.4.

4. **Transformation**
   After estimating the homography parameters, the sensed image is transformed. The transformation includes translation, rotation and scaling.

Figure 1.1 shows the whole process of registration. The yellow dots in these two images represent the feature points. Using similarity measures on these feature points, the feature correspondence is established which is described by the green lines in this figure. Then homography between images is estimated by mapping function. The parameters for rotation, translation and scaling are found. At last, the sensed image is transformed by this homography parameters, such as rotation, translation and scaling.

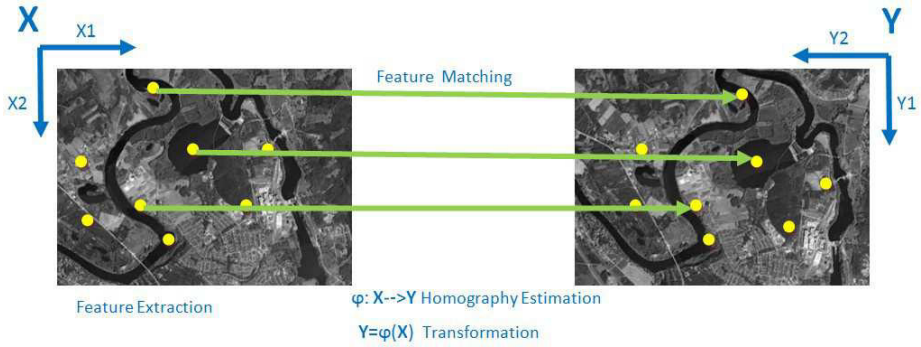


Figure 1.1: The whole process of registration

Each registration step has its own unique challenge. At the beginning, it should think about what kind of features are suitable for the given task. The features should be stable, distinctive objects, not sensitive to image degradation or deformation. The feature sets are expected to have sufficient common elements in both reference and sensed images.

In the next step, incorrect feature extraction or image degradation will lead to mistakes. The feature descriptor should be invariant to image degradation.

Different feature descriptors and similarity measures are used for feature matching. The similar feature points are found by one of similar measurements.

In the homography estimation step, the mapping function is chosen by prior information concerning the acquisition process and expected image degradations. Without prior information the mapping function should be flexible and robust to all the possible degradations. In this step the accuracy of the feature extraction, the reliability of homography estimation and the acceptable approximation error should be considered.

Finally, the transformation method depends on the trade-off between the required accuracy of the interpolation and the required computational time.

1.2 Aerial Imaging

Aerial imaging and mapping present a "bird's-eye" view of the earth. While aerial images are not maps.

Maps are orthogonal representations of the earth's surface. They are directionally and geometrically accurate. Aerial imaging displays a high degree of radial distortion. It means the topography is distorted. So the measurements made from an aerial image is not accurate. However, aerial images are powerful tools for studying the earth's environment. Figure 1.2 shows a sample of an aerial image.

Due to most geographic information system's (also called GIS, which is a system designed to deal with all type of geographical data) can correct for radial distortion, aerial images are very high quality data sources for many types of projects, especially those that require spatial data from the same location at periodic intervals over a length of time.

Arial images are different from "regular" photo images in at least three ways. They portray objects from an overhead (and unfamiliar) position, often record infrared wavelengths and are taken at scales most people are unaccustomed to seeing

Here the advantages of Aerial Imaging has been described over Ground-Based Observation. It offers an improved vantage point, and provides a permanent recording. Aerial images have broader spectral sensitivity than the human eye, and have better spatial resolution and geometric fidelity than many ground-based sensing methods.

Figure 1.2: Aerial image.

1.3 Problem Statement

The common registration methods are firstly, taking control points over the images to correct the geometric errors such as translation, rotation and scaling, and secondly, establishing the correspondence model between images. It does not matter which registration method and particular image is used, the most important thing is an estimate how accurate the registration is. However, each step of registration will influence its final result.

Phase correlation, based on the translation property of the Fourier transform, computes the FFT ratio instead of any control points. The sharp peak of the inverse of this ratio represents the highest correlation. Before phase correlation, the reference and sensed images are transformed from the Cartesian coordinates to the logploar image and then phase correlation computes the similar ratio, rotation and scaling errors which are represented as shift. The correspondence is establish by using these parameters.

The previous registration methods use the whole image to do logpolar transformation and phase correlation. However, the common log-polar transformation for image registration is to crop a square patch in both the reference image and the sensed image. After the image has been transformed from Cartesian coordinates to Log-Polar coordinates, black semicircles appear on the border of the image, overlapping to form sharp edges / corners in between. Phase correlation focuses on these corners at the black semicircles. There is no information at

these black semicircles, which influences the accuracy of the registration. Figure 1.3 shows a sample of previous registration methods.

Figure 1.3: The previous registration method.

1.4 Proposed Solution

Aerial image registration requires a high degree of precision. In order to improve the accuracy of feature-based registration, this project proposes a novel Log-Polar Transform (LPT) based image registration. Instead of using the whole image in the conventional method, feature points are used in this project, which reduces the computational time. Furthermore, the black semicircles need to be avoided. For rotation invariant, it is not important how the image patch is rotated. The key is focusing on the feature points. Image registration with FFT always do the FFT first.

The proposed process of this project contains four steps. First, feature points are selected in both the reference image and the sensed image with corner detector (Harris or SIFT). Secondly, image patches are created using feature point positions as centers. Each point is a center point of LPT, so circular image patches are cropped choosing feature point as center. The radius of the circle can be changed, then the circular images are transformed to Log-Polar coordinates. Next, the LPT images are dealt with using phase correlation. Figure 1.4 shows the proposed solution in this project, which removes the black semicircles in success after Log-Polar Transformation.

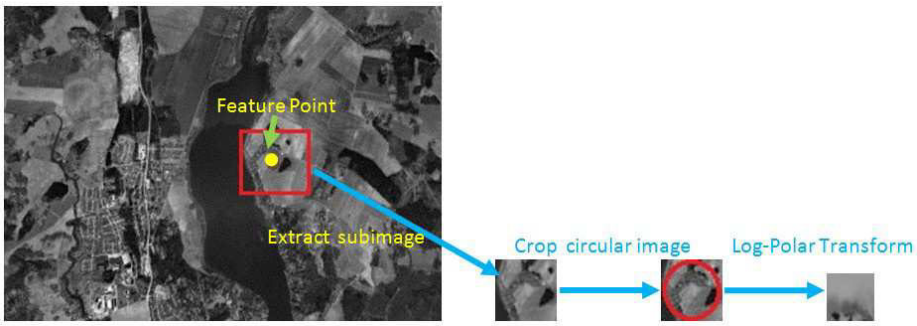


Figure 1.4: A proposed solution.

1.5 Overview of the Thesis

This thesis is organized as follows: Each chapter has its own topic, which builds on the work discussed in the earlier chapter. Chapter 2 describes the theoretical background and reviews the existing techniques of registration. Chapter 3 presents the outline of our novel registration system. Chapter 4 gives the experimental results and Chapter 5 analyzes and discusses these results. Chapter 6 provides the conclusion and Chapter 7 lists further extension work.

Chapter 2

Theoretical Background

2.1 Literature Review

This section contains an overview of the previously published, related image registration literature.

Barbara Zitov and Jan Flusser [33] present the classical and up-to-date image registration methods. All the methods can be classified into two categories: area-based methods and feature-based methods. Area-based methods do not extract features, so they don't have the feature extraction step of image registration. These methods put emphasis on feature matching. Feature-based methods extract salient structures (also called features) in the images. Region features (lakes, fields and forests) are detected by means of segmentation methods [23]. Line features(roads, rovers and region boundaries) are extracted by standard edge detection. Point features (region corners, line intersections and points on curves with high curvatures) are widely used for registration.

Tuytelaars and Krystian [30] survey invariant interest point detectors, discuss the properties of the local feature and analyze their strengths and weaknesses. The Harris Corner detector is proposed by Harris and Stephens [11], which is based on the second moment matrix and can achieve sub-pixel precision. It is also proven to be the most informative and repeatable [26] [27]. Although it takes long time for computation, this detector is widely used in practice. Because the Harris corner detector is influenced by image scale, it could not match images in different sizes.

The Smallest Univalue Segment Assimilation Nucleus (also called SUSAN) corner detector is introduced by Smith and Brady [29]. For each pixel in the image, a circular neighborhood of fixed radius is considered around it. The

SUSAN corners are detected by segmenting this neighborhood into two regions: "similar"region and "dissimilar"region. The corners are located at the place where the "similar"region reaches a local minimum below a certain threshold. The SUSAN detector is more efficient than Harris but also more sensitive to noise.

Tuytelaars and Krystian [30] also discuss Harris-Laplace and Harris-Affine. Harris-Laplace starts with a multiscale Harris corner detector to determine the location of the local features. The Laplacian operator is used to select scale. The select scaled measures the scale at the place where there is maximum similarity between the local image structures and the feature detection operator. Harris-Affine obtain affine invariant corners, which detect the initial region with the Harris-Laplace detector.

David G. Lowe [17] [18] [19] uses a descriptor called Scale Invariant Feature Transform (SIFT). It extracts the invariant feature points in a pyramidal Difference of Gaussians (DOG) and can be invariant to image scaling, translation and rotation, and partially invariant to illumination changes and affine or 3D projection.

There are some extensions of SIFT descriptors, such as PCA-SIFT, GLOH, SURF and ASIFT. The descriptor of SIFT uses 128 dimensions, whereas less are used in PCA-SIFT. PCA-SIFT [16] uses Principal Components Analysis (PCA) to normalize the gradient patch, instead of using SIFT's smoothed weighted histograms. The dimension of descriptor infulences the running time for interesting point matching. The less dimension is required for fast matching. PCA-SIFT yields a 36 dimensions, so it is faster for matching. But it is sensitive to registration localization error. Herbert Bay, et al.[4] use a descriptor called Speed-Up Robust Features (SURF). Compared with SIFT which builds pyramids into different resolutions, SURF uses another method to extract feature points. It depends on integral images for image convolutions and a Hessian matrix-based measure for detector. Once the integral images have been computed , this descriptor runs fast, taking only three additions and four memory accesses to calculate the sum of intensities inside a rectangular region of any size.

Michael Grabner et al.[4] propose a fast SIFT descriptor which calculates the integral images. It is based on orientation histograms. Compared with SIFT, it computes the scale space by Difference of Mean(DoM) images instead of Difference of Gaussians(DoG). When the overlap error tolerance is large enough, the fast SIFT descriptor works better than the original SIFT. Faraj Alhwarin et al. [9] improve SIFT feature matching and divide the extracted features into some sub-collections before the two images are matched. This method reduces the running time for matching images. Luo Juan [14] highlighted that SIFT is slowest and not as good at illumination changes. SURF is fast but not stable to rotation and illumination changes. PCA-SIFT should improve on blur

performance. The methods are selected depending on the application.

Krystian Mikolajczyk and Cordelia Schmid[22] evaluate ten different descriptors for different types of interest regions. By comparing the ranking of the descriptor in affine transformations, scale changes, rotation, blur, jpeg compression and illumination changes, they point out gradient location and orientation histogram (GLOH) can provide the better results than SIFT. The SIFT based descriptors perform better than any other region detectors.

In the matching step, the area-based methods are Cross-correlation (CC), Sequential Similarity Detection Algorithm (SSDA), Mutual Information (MI), Phase Correlation and Optimization methods.

CC measures the similarity by computed for window pairs from the reference and sensed images and searched its maximum. CC matches image intensities directly without structural analysis. It is sensitive to changes in intensity, which depends on noise, illumination and different types of sensor. SSDA [3] accumulates the sum of the absolute differences in image intensity values. It makes use of a threshold to judge the candidate pair of windows from the reference and the sensed images. This method is faster but less accurate than CC. Which are both correlation methods. The main disadvantages of correlation-like method are the high computational time required and the flatness of the similarity measure maxima.

The phase correlation [6] presents correlation matching between the two images in the frequency domain. Compared with cross-correlation, it performs well with noise and illumination changes. If some changes in rotation or scale exist between the two images, the correlation methods will fail.

Log-Polar Transform(LPT) [2] [15] [20] [25] [34] [32] is widely used for its scale and rotation invariant properties. Rotation and scale change in Cartesian coordinate are described as translation in the log-polar domain. In the log-polar domain the image pixels close to the center are oversampled, but the image pixels far away from the center are undersampled.

Fourier-Mellin combine the phase correlation technique with the Log-Polar Transformation. It describes rotation and scale as translations in parameter space[7]. Anuta [1] recommends to compute the correlation in the frequency domain.

Mutual Information (MI) represents the leading technique in multimodal registration, especially in medical image registration. It works with the intensities on the whole image data and measures how much information one random variable tells about another.

Dubrofsky [8] discusses homography estimation in various situations. Robust estimation techniques deal with outliers. Hongshi and Jian Guo[31] present a phase correlation based sub-pixel feature matching technique.

Hast [13] presents the process of making a spatial atlas with historic images. His aim is rebuild a virtual globe, similar to Google Earth (GE) but oriented towards the management of the temporal variable.

2.2 Feature Extraction

2.2.1 Preprocessing

This section discusses some preprocessing methods before explaining the image registration. One possible problem is that the aerial image might not be in good condition for registration. For example, old aerial photos may appear yellow due to decay or general wear and tear. Additionally, a better stitching result will not be achieved if the two aerial images have different illuminations, even after registration. Another problem is that it takes a long time to deal with the high resolution images. For these reasons the aerial image should be preprocessed. Then they can be used for feature extraction. According to the above problems, the aerial image is dealt with the following method.

- **Modify Intensity**
  Although the intensity does not influence the accuracy of phase correlation in the proposed registration method, it might get a bad stitching result after registration. This then would produce a visible seam in the joining of images. In order to implement a better blending operation, modifying the intensity of the registration image is suggested. For example, changing the brightness of images makes these images have same intensity value. This operation can also be done after registration, but it is recommended to be done prior to assure a better result.

- **Remove Noise**
  Noise in the image will lead to inaccurate results. In spite of the feature descriptor which are selected to cope well with noise, it is still better to remove noise using a smoothing operator. Guassian filter and Median filter are often used to blur the image and remove noise.

2.2.2 Feature Extraction

This section discusses the properties of the local feature. Next, it explains the basic idea of *Harris corner detection*, a derivatives-based approach. It then move on to *Scale Invariant Feature Transform (SIFT)*.

Good features should include the following properties [22] :

- *Repeatability*: Given two images of the same scene with different viewpoints, a high percentage of the feature detected on the scene in both image should be found in both images.
- *Distinctiveness/ informativeness*: The features can be distinguished and matched using the different intensity patterns.
- *Locality*: It reduces the probability of occlusion if the feature are local.
- *Quantity*: The adequate number of features should be detected, so that the small object will contain a reasonable number of features.
- *Accuracy*: The features should express the location and shape accurately.
- *Efficiency*: The feature detected in a new image should satisfy the time-critical applications.

Repeatability is the most important one among these properties. It can be achieved in two different way: *invariance* or *robustness*.

Invariance: The preferred method of feature detection is not affected by math transformations when we expect a large deformation.

Robustness: The deformation such as image noise, compression artifacts, and blur influences the accuracy of detection. The feature extraction method is expected to be less sensitive to deformation, so that the accuracy will not change a lot.

Harris Corner Detector

The Harris Corner Detector is one of the simplest corner detector algorithm, proposed by *Chris Harris* and *Mike Stephens*. It is based on the *second moment matrix*, called the auto- correlation matrix. The main idea is to find the interest points called corners where the surrounding neighborhood shows edges in more than one direction.

The Harris Corner Detector algorithm has the following steps:

1. Define a matrix M for each pixel (x,y) in the image domain, as the positive semi-definite, symmetric matrix

$$M = \begin{bmatrix} I_x \\ I_y \end{bmatrix} \begin{bmatrix} I_x & I_y \end{bmatrix} = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{2.1}$$

2. Let w be a weight matrix(a Gaussian filter), the convolution

$$\overline{M} = w * M \tag{2.2}$$

The width of w determines the interesting region around x. The result $\overline{M}$ is auto- correlation matrix, sometimes called Harris matrix. It gives a averaging of M over the neighborhood pixels.

3. For each pixel (x,y), construct an indicator function, calculate the cornerness(x,y).

$$cornerness(x, y) = det(\overline{M}) - k(trace(\overline{M}))^2 \tag{2.3}$$

$$det(\overline{M}) = \lambda_1 \lambda_2 \tag{2.4}$$

$$trace(\overline{M}))^2 = \lambda_1 + \lambda_2 \tag{2.5}$$

k is a constant. A typical value for k is 0.04.

3. Set all cornerness(x,y) below a threshold T to zero. Threshold T controls the number of corners.

SIFT

The scale invariant feature transform (SIFT) algorithm was developed by David G. Lowe [17] [18] [19]. The SIFT features are invariant to image scale and rotation and cope well with changes in illumination and 3D viewpoint. Because of the highly distinctive property of the SIFT feature, a single feature can be correctly matched with high probability against a large number of features[19].

SIFT feature detection works in four consecutive steps.

- scale-space extrema detection
- keypoint localization
- orientation assignment
- keypoint descriptor

Details of the above steps are given in the following:

1. In order to find out the potential feature points which are invariant to scale and orientation. This step is implemented by using the extrema of Difference of Gaussians(DOG) to search over all scales and image locations.

   First, it establishes the scale space with Gaussian function. The scale space can be written as

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{2.6}$$

Where * is the convolution operation, I(x,y) is an input image and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{2.7}$$

is the Gaussian function in 2D situation.

While σ needs to be assigned a different value to construct n scaled image in its first scale space. The DOG image is obtained by the Difference of Gaussian function convolved with the image I(x,y). It can be described as

$$D(x, y, \sigma) = ((L(x, y, k\sigma) - L(x, y, \sigma)) * I(x, y) \tag{2.8}$$

which is equal to

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{2.9}$$

where k=$\sqrt{2}$.
The size of each octave scale space is half the previous one. As shown in Figure 2.1, on the left a set of scale space images are produced by the initial image convolved with Gaussian repeatedly. On the right, adjacent Gaussian images are substracted to produce the difference of Gaussian(DOG).

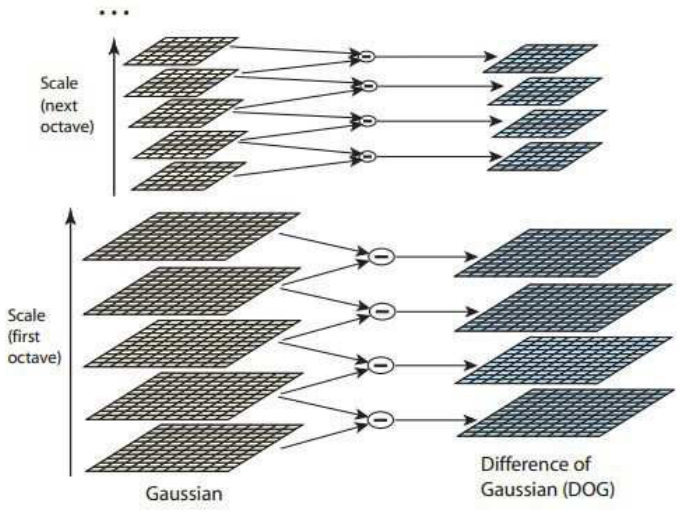


Figure 2.1: Construction of DOG image(Image source:[19]).

Secondly, it detects the local extrema(maxima or minima) of the DOG image. Looking at Figure 2.2, each pixel is compared against 26 neighbour points (8+2×9=26, 8 neighbours in the current scale space image, 9 neighbours in the scale above and 9 neighbours in the scale below). The pixel is selected only if its value larger or smaller than all of these neighbors. In the 2D situation, we need to construct 4 different scale images. Because it needs to compare 3 DOG images in each octave[19].

2. The local maxima and minima of DOG are good candidates for feature points. But the local extrema with low contrast or are poorly localized such as correspond to edges should be removed. The main reason is that they are sensitive to noise. Taylor expansion of the value at the local extrema can find the intensity. For stability, it should reject the points whose locations are along the edge. As the same approach of Harris Corner Detector, it computes Hessian matrix and checks that the ratio of principal curvatures is under some threshold, r. We need to check

$$\frac{trace(M)^2}{det(M)} < \frac{(r+1)^2}{r} \tag{2.10}$$

3. The keypoint descriptor can tolerate image rotation by being assigned an orientation . The following method can provide the most stable results. For each image sample, *L(x,y)* (Gaussian smoothed image), the gradient of each key point should be calculated. The magnitude *m(x,y)*and orientation $\theta(x,y)$ of the gradient is computed respectively as

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \tag{2.11}$$

$$\theta(x,y) = tan^{-1}\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \tag{2.12}$$

According to Lowe's suggestion [19], an orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint, which has 36 bins covering the 360 degree range of orientations. After the highest peak in the histogram is detected, any other local peak with 80% of the peak is used to also create a keypoint with that orientation. There will be multiple keypoints created at the same location and scale but with a different orientation because of locations with multiple peaks of similar magnitude.

4. The previous steps define an image location, scale and orientation of each keypoint. This step is to compute a descriptor which is highly distinctive and invariant to 3D viewpoint or changes in illumination. Figure 2.3

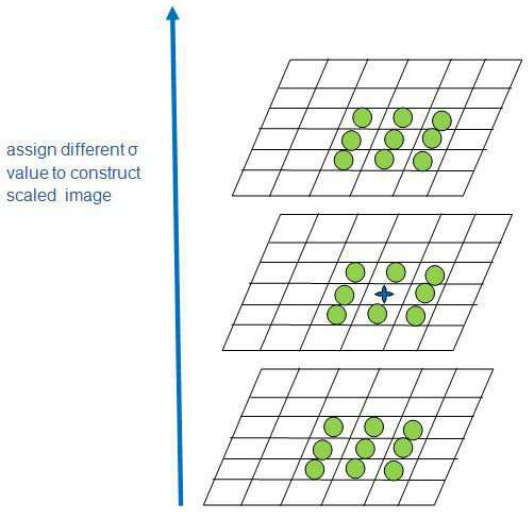


Figure 2.2: Maxima and minima of the difference-of-Gaussian images

illustrates how to create a keypoint descriptor. Here a 16×16 Gaussian window is selected around the keypoint location and broken it into 4×4 subwindows. In each 4×4 subwindow, it calculates the magnitude *m(x,y)* and orientation *θ(x,y)* of the gradient at each point. Next, these orientations are put into an 8 bin histogram with contains 8 features in each subwindow. In this case, the 16×16 window is used to create a keypoint, so that each keypoint has 128 feature vectors(16×8=128).

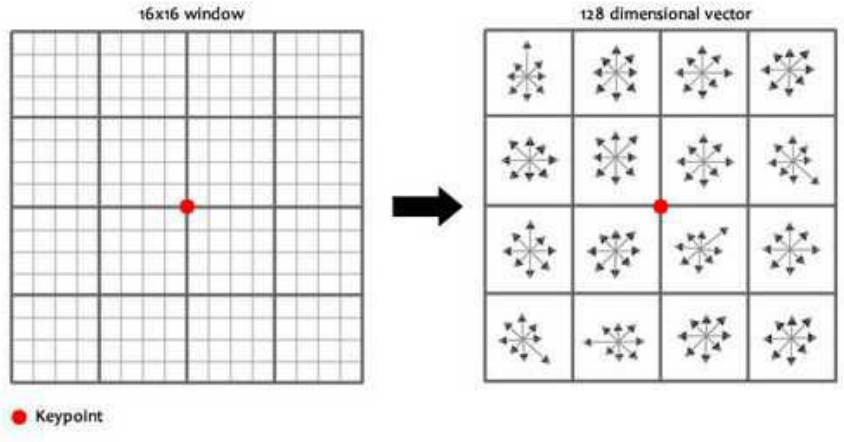


Figure 2.3: Creation of keypoint descriptor*(*Image source:[28]).

2.3 Feature Matching

Based on the feature points extracted from the previous chapter, feature matching can now begin. Original feature matching process establishes correspondence between the reference image and the sensed image. The novel method is matching the correspondence circular images patches in both images. As explained in section 4, we will get a lot of key point region images and their descriptors for each image.

2.3.1 Log-Polar Transform

Log-Polar Transform (LPT) is a nonlinear sampling method which converts an image from the Cartesian coordinates I(x,y) to the logpolar coordinates LPT(ρ, θ). The mapping function is given by

$$\rho = log_{base}\sqrt{(x - x_c)^2 + (y - y_c)^2} \tag{2.13}$$

$$\theta = tan^{-1}\frac{y - y_c}{x - x_c} \tag{2.14}$$

where
(x_c, y_c) is the center pixel of the transformation in the Cartesian coordinates,
(x, y) is the sampling pixel in the Cartesian coordinates,
(ρ, θ) is the log-radius and the angular position in the log-polar coordinates.

As shown in figure 2.4 , from the center to the furthest circumference the distance between two successive sampling points in the radius direction increase exponentially. In the angular direction, the circumference is sampled with the same number of samples. So image pixels close to the center are oversampled while image pixels far away from the center are undersampled or missed. In the log-polar coordinates, the row represents the log-radius while the column represents the angular direction.

The benefit of using log-polar coordinates instead of Cartesian coordinates is that any rotation and scale in the Cartesian coordinates are represented as translation in the log-polar coordinates. So LPT based image registration can deal with changes in rotation and scale. The common method transforms a square image to log-polar coordinates. Due to undersampling or exempt information, the LPT transformed image appears with black semicircle shape regions.

A second benefit of log-polar coordinates is the space variant property. The sampling resolution in the Cartesian coordinates is higher around the center point and descreases exponentially as it gets far away from the center. So the center of the LPT becomes more vital than the surrounding areas which are far away form the feature point center.

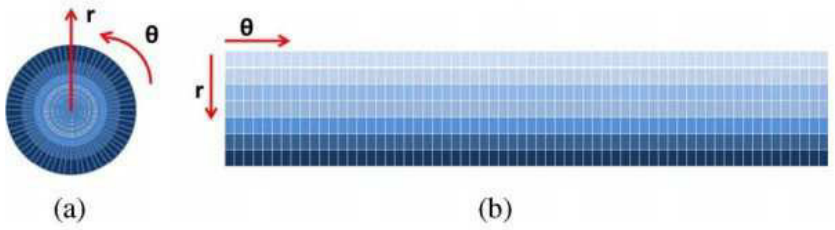


Figure 2.4: LPT mapping: (a) LPT sampling in the Cartesian Coordinates, (b) the sample result in the angular and log-radius directions (Image source:[21]).

Based on one feature point as the center of the circle, our novel method extracts a circle image patch from Cartesian coordinates to the log-polar coordinate. It can therefore avoid the black semicircle shape regions . Without the noise regions, the next step phase correlation can become more accurate.

2.3.2 Fourier Method

The frequency domain is a space in which each image value at image position F represents the amount that the intensity values in image I vary over a specific distance related to F. In the frequency domain, changes in image position correspond to changes in the spatial frequency, (or the rate at which image intensity values) are changing in the spatial domain image I.

There are some reasons to explain why the frequency domain is so interesting. On the one hand, it can make explicit periodic relationships in the spatial domain. On the other hand, some image processing operators are more efficient when they are applied in the frequency domain.

The Fourier Transform method converts images from the spatial domain into the frequency domain. The inverse Fourier Transform method is used to convert images from the frequency domain into the spatial domain.

Fast Fourier Transform

The Fourier Transform is a technique which is used to convert images from the spatial domain into the frequency domain and decompose an image into its sine and cosine components. In the Fourier domain image, each point represents a particular frequency contained in the spatial domain image.

The main idea of Discrete Fourier Transform is that F(u,v) in the frequency domain represents the image f(x,y) of size M x N in the spatial

domain. It is the sampled Fourier Transform , so it does not contain all frequencies forming an image. But the set of samples is large enough to fully describe the spatial domain image.

The two-dimensional discrete Fourier transform (DFT) is computed as follows:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi(ux/_M + vy/_N)} \tag{2.15}$$

The concept behind the Fourier transform is using a sum of sine and cosine waves of different frequencies to construct any waveform. In equation (2.15) the exponential can be expanded into sines and cosines with the variables u and v determining these frequencies.

The equation for the inverse of the above discrete Fourier transform is

$$f(x,y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{j2\pi(ux/_M + vy/_N)} \tag{2.16}$$

So, if F(u,v) is known, using the equation(2.16) of the inverse, discrete Fourier transform the corresponding image f(x,y) can be gotten.

Here are some notes about the discrete Fourier transform:

- The values of the Fourier transform are complex, meaning they have real and imaginary parts.
- We visually analyze a Fourier transform by computing a Fourier spectrum (the magnitude of F(u,v)) and display it as an image. The Fourier spectrum is symmetric about the origin
- The fast Fourier transform (FFT) is a fast algorithm for computing the discrete Fourier transform.

Phase Correlation

This subsection describes the phase correlation method and outline its vital properties.

The Phase Correlation method is proposed for the registration of translated images, which is based on the Fourier Shift property[5]. It transforms shift as phase difference in the Fourier domain.

The Phase Correlation algorithm can be described as follows:

(a) Give two input images g_a and g_b, then apply a window function (e.g., a Hamming window) on both images to reduce edge effects. Next, calculate the discrete 2D Fourier transform of both images.

$$G_a = F\{g_a\},\ G_b = F\{g_b\}$$

The experimental results show that the images without a window function get better results in this case, so we can ignore this step.

(b) Normalize the cross-power spectrum. It is given by

$$R = \frac{G_a G_b^*}{\mid G_a G_b^* \mid}$$

where * is the complex conjugate.

(c) Apply the inverse Fourier transform to get the normalized cross-correlation.

$$r = F^{-1}\{R\}$$

(d) Look for the location of the peak in its inverse Fourier transform r.

$$(\Delta x, \Delta y) = argmax_{(x,y)}\{r\}$$

Properties of The Phase Correlation algorithm

- The most important property of this algorithm is the value of the distinct, sharp peak showing the accuracy of correlation. Figure 2.5 is an example of two displaced aerial images. The phase correlation yield is a sharp peak which represents a high correlation. The standard cross correlation produces some broad peaks while the maximum peak (also called the main peak) is not located at exactly the right point.
- Another vital property is its resistance to disturbances in illuminate and offsets in average intensity.

2.3.3 Correlationlike Method

Correlation between two image patches, $I_1(x)$ and $I_2(x)$, is defined as:

$$c(I_1, I_2) = \sum_x f(I_1(X), I_2(x)) \tag{2.17}$$

where the function f depends on the correlation method. The sum is calculated by all positions x in the image patches. For cross correlation $f(I_1, I_2) = I_1 I_2$, and then $c(I_1, I_2) = I_1 \cdot I_2$ with '$\cdot$ 'presenting the scalar

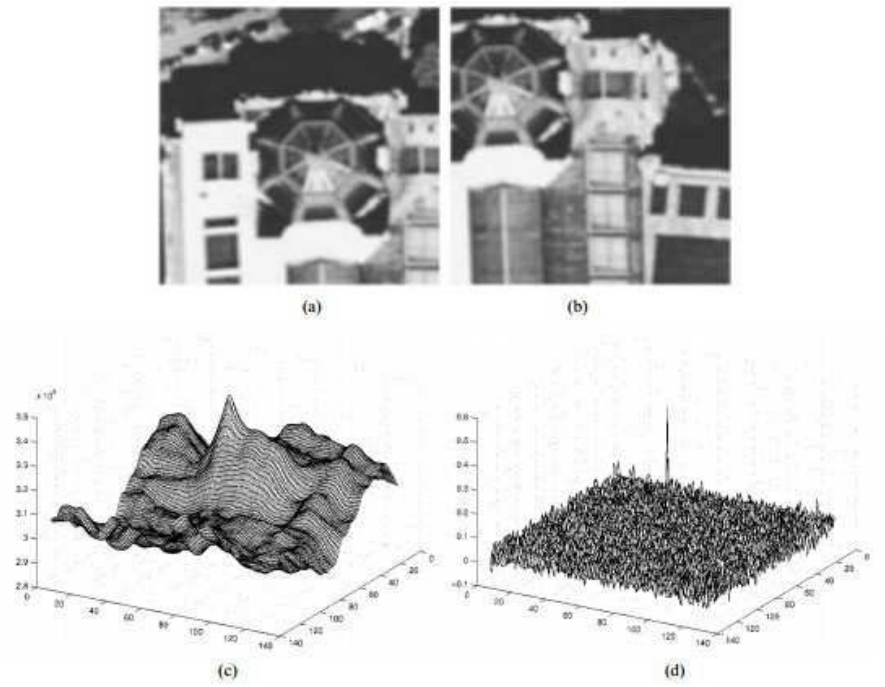


Figure 2.5: Phase correlation and standard cross-correlation:(a) and (b) Aerial image of Paris with displacements along both axes, (c) standard cross-correlation and (d) phase correlation *(Image source:[10]).*

product. The larger the value of $c(I_1, I_2)$, the more similar the patches $I_1(x)$ and $I_2(x)$ are.

The classical representation of the area-based method is normalized cross correlation. It is given by

$$ncc(I_1, I_2) = \frac{1}{n-1} \sum_x \frac{I_1(x) - \mu_1}{\sigma_1} \cdot \frac{I_2(x) - \mu_2}{\sigma_2} \tag{2.18}$$

where n is the number of pixels in a patch, μ_1 and μ_2 are the mean intensities, σ_1 and σ_2 are the standard deviations in each patch. The method copes well with the image brightness by subtracting the mean and scaling with the standard deviation. If it requires the subpixel accuracy of the registration, the interpolation of the normalized cross correlation measure values need to be used. This method can exactly align translation images. It can also be applied in slight rotation images and slight scaling images.

2.3.4 Removing False Matches

Symmetry Test

In the Symmetry Test, the matches are computed using the symmetrical (two-sided) matching conditions to remove the false match (i.e. the match pair could not pass the symmetry test in both directions) and keep the accurate match.

2.4 Homography Estimation

Homography is mapping points and lines in one plane to another plane. The common algorithm uses point feature correspondences to estimate homographies between images.

According to the homography definition from Hartley and Zisserman [12], 3×3 homography matrix H is suffcient to calculate the homography. The homography matrix H has 8 independent degrees of freedom.

2.4.1 Algorithms for Homography Estimation

Basic DLT algorithm

The Direct Linear Transform (DLT) algorithm is an algorithm for computing the homography matrix H if it gives four or more point correspondences. The 2D points are mapping in homogeneous coordinates, the relationship between corresponding points x and x' can be described as

$$x' = Hx \tag{2.19}$$

or

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} \tag{2.20}$$

In homogeneous coordinates the points are scaled, so the 2D point can be expressed as x=[x,y,w]=[αx,αy,αw] = [$x/_w$,$y/_w$,1]. To get a unique identification of the image coordinates x,y, the points are often normalized with w=1. In order to obtain the correct result, Hartley and Zisserman [12] propose a normalization step on DLT.

Different Cost Functions

Minimizing a suitable cost function should be considered to solve for a homography when there are more than 4 points of correspondence available. Some of these cost functions are used in practice:

- **Algebraic distance**: The simplest cost function is to minimize the algebraic distance using Singular Value decomposition (SVD) analysis. Although it is cheap, the quantity being minimized is not geometrically meaningful.

- **Geometric distance**: This measures the Euclidian image distance between the point where the homography maps and the correspondence point. An iterative method is required to minimize the cost function. Although the results are more accurate, iterative techniques have disadvantage when compared with minimizing Algebraic distance. The method is slower, risk not converging and so on.

- **Reprojection error**: The goal of the reprojection error cost function is to make a correction for each correspondence. Minimizing this function is more complicated than Geometric distance.

- **Sampson error**: On one hand, algebraic cost function is cheap to compute but does not provide geometrically meaningful results. On the other hand, the geometric and reprojection error cost function can give more accurate results but their iterative methods are more complex. The Sampson error cost function lies in between two extremes which are computation cost and providing a close approximation to represented error.

2.4.2 Robust Estimation

The homography estimation algorithm requires a number of correspondences(at least 4 correspondences). However, some situations may arise where the two features in the images do not correspond to the same real feature. Therefore we need to distinguish inlier [1] and outlier [2] correspondence. Only inlier matches can be used for homography estimation. In this section we will discuss two homography estimation methods.

RANSAC

RANSAC (Random Sample Consensus) is the most popular robust homography estimation method [8]. The method can be described like this. First, it selects 4 correspondences at random and computes a homography matrix H from these 4 correspondences in each iteration. Secondly, it classifies the other correspondences as inliners or outliers depending on its concurrence with H. Next it repeats this process in all iterations and then chooses the iteration which contains the largest number of inliers. The homography matrix H in that iteration is considered as the best

[1] inlier, the true matching correspondence, is used for homography estimation
[2] outlier, the wrong matching correspondence, is not considered for homography estimation

transformation model.

When RANSAC classifies correspondences as inliers or outliers, a distance threshold t between x' and Hx needs to be assigned. If the distance $> t$, the point is considered an outlier. We also need to decide the number of iterations for running RANSAC. Hartley and Zisserman [12] suggest the formula that

$$N = \log(1-p)/\log(1-(1-\epsilon)^s) \tag{2.21}$$

where ϵ is the probability of outliers and s is the number of correspondences used in each iteration ($s = 4$ in this case).

Least Median of Squares Regression

A distance threshold is chosen to find out outliers and discard them in RANSAC . It deals with the sum of squared difference algorithms. However, the Algebraic distance version of DLT is not very robust with outliers.

The Least Median of Squares (LMS) estimation method replaces the sum of squared residuals. Peter J.Rousseeuw [24] states that LMS works very well if there are less than 50% outliers and it does not need to set any parameters such as the threshold or the probability of outliers. The major disadvantage is that LMS could not deal with the outliers, which are more than half the data.

Chapter 3

System Description

This chapter introduces the novel registration system. The purpose of the project is to analyze the transformation relationship between the reference image and the sensed image. For obtaining a satisfying final result, preprocessing work is usually done to remove noise and modify intensity after selecting interesting aerial images.

At the beginning, the corner detector method is used to obtain feature points in the reference image and the sensed image. Circle image patches are created using these feature points as centers. In order to deal with rotation situation, circle shaped image patches are used. It can avoid the black semicircle shaped noise after log-polar transformation. Next, we transform all the image patches from the Cartesian coordinates I(x,y) to the logpolar coordinates LPT(ρ, θ). Any rotation and scale in the Cartesian coordinates are represented as translation in the log-polar coordinates. Then Fast Fourier Transform is used to convert log-polar based images from the spatial domain into the frequency domain. Based on the Fourier Shift property, the Phase Correlation method is proposed for the registration of translated images. Symmetry test is used for removing false matches. A good matching result can give accurate transformation parameters. At last, RANSAC is used for homography estimation. Figure 3.1 shows the whole registration system proposed in this thesis.

Some parameters will influence the accuracy of this novel registration system. The radius of circle image patches and the size of input images of FFT are the important parameters in this project. As shown in Figure 3.3, the radius of circle image patches decides the size of input image of LPT. It also points out the size of input images of FFT affects the FFT. The output image of the LPT is the same one for input image of FFT. In order to get high accuracy, the Area of the input image of LPT should be bigger than the area of input image of FFT.

One detail about the radius of circle image patches should be mentioned.

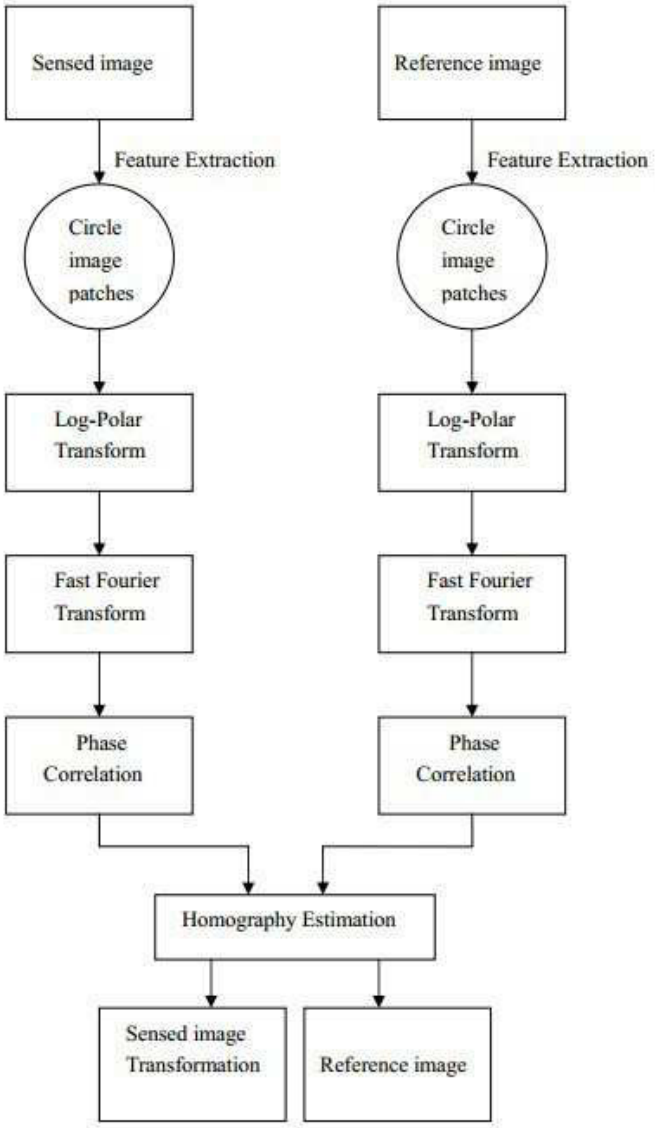


Figure 3.1: Rigistration System Overview.

The pixel of the center point is removed for avoiding oversample in LPT. Figure 3.3 shows this detail when the size of subimage square is $m \times m$, the

radius of the Largest cirlce image patch inside it is defined as $m = 2 \times radius + 1$.

The radius of circle image patches will be tested, accuracy of registration with a set of radius value compared the best value for the radius investigated. The size of input image of LPT will be tested also. Then two methods are compared—-first Log-polar then FFT or first FFT then Log-polar.

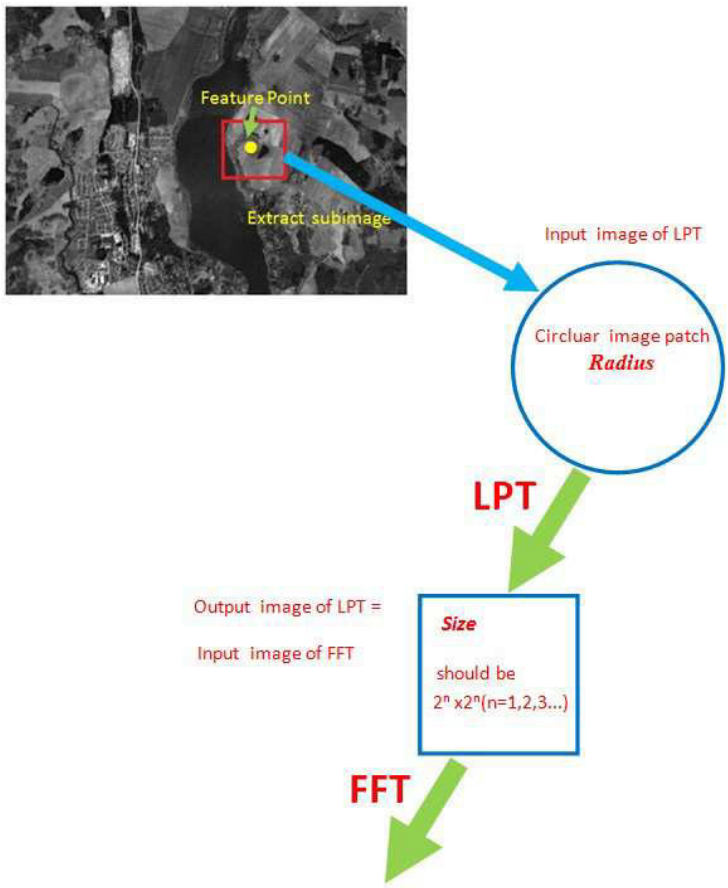


Figure 3.2: The parameters that affect registration are ***radius*** and ***size***.

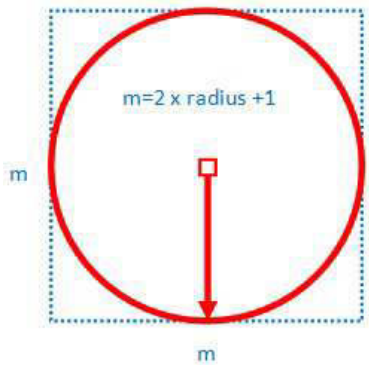


Figure 3.3: The radius of circular image patch.

Chapter 4

Experiments and Results

This chapter presents experimental results on the proposed algorithm. The algorithm is developed using *Python 2.7*, used Intel *OpenCV (Open Source Computer Vision)* library version 2.4.3 to test the performance of the homography and carried out these tests in one computer with the following configurations:

- System type: Windows 7 Professional 64-bit Operating System
- Processor: Intel Core i5 2.80 GHz
- Memory: 8.00 GB

All of the images that have been tested in this project are screenshots from Google Earth (Product version 7.0.3.8542). The resolution of these images are 418*320.

Regardless of the used registration method and the particular images, the most important thing for the user is an estimate how accurate the registration is. The accuracy evaluation is not a normal problem, because the errors can be dragged into each of stages in the registration process and it is difficult to distinguish between registration inaccuracies and physical differences in the image contents.

For accuracy evaluation of these experimental results, the homography with RANSAC was estimated ten times and the average number of inliers was calculated. In each actual experiment, same results were gotten in ten times. It proved the novel method is stable and reliable.

The novel registration method is called LPT_PC (Log-Polar Transform and Phase correlation) in this project. At first, the experiment tested different parameters and selected the optimal parameters for the novel algorithm. Then it compared the novel algorithm LPT_PC using SIFT selecting feature points with the original SIFT (developed by David G.Lowe).

Next, it tested the LPT_PC (Log-Polar Transform and Phase correlation) and LPT_CC (Log-Polar Transform and Cross Correlation) using Harris selecting feature points in both methods. The novel method will now be evaluate based on the number of inliers.

4.1 Results of Optimal Parameter Selection

In this section the optimal parameters for the novel algrithm will be discussed. Different parameters for the radius of the circle image patch in LPT will also be compared, as well as the input size of FFT. Then these optimal parameters will be used to registration.

Here 400 feature points are extracted with SIFT in the reference images and the sensed images. Because the interesting square is extracted from the original image, using feature points as centers, when the square is 41x41, the biggest radius should be 20. The left image in Figure 4.1 shows this principle. In order to get high precision, the input size of FFT should be smaller than the size of the interesting square. Furthermore, the input size of FFT should be $2^n (n = 1, 2, 3, 4....)$. When the square is 41x41, the biggest input size of FFT is 32x32. The input size of FFT — 64x64 needs to be tested, so the size of the the interesting square should be bigger than 64x64. Here 71x71 is chosen for the input size of FFT — 64x64. The right image in Figure 4.1 shows the biggest the radius in this case. In these tests, the radius of the circle image patch in LPT are 10, 12, 14, 16, 18, 20. The input sizes of FFT are 8x8, 16x16, 32x32, 64x64. 10 sets of images have been tested. Figure 4.2 is one sample of test results. Other test results are in Appendix A.

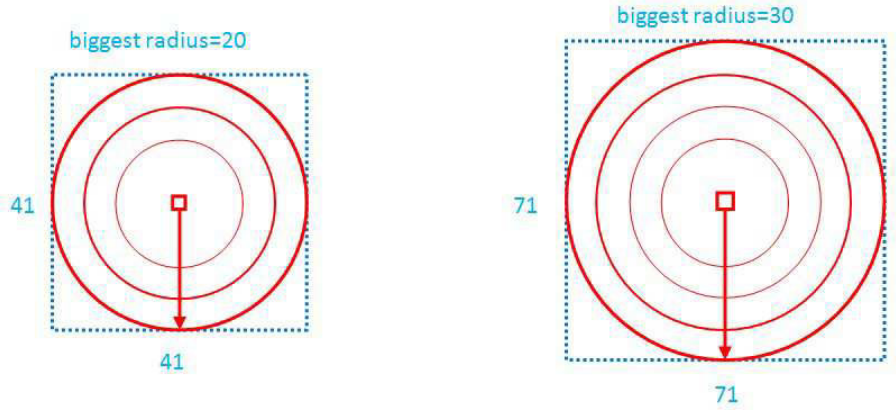


Figure 4.1: The principle of selecting parameters.

The chart presented in Figure 4.2 shows the optimal parameters —— the

radius for LPT is 20, the input size of FFT is 32x32. In this case, the highest number of the inliers is 239 for this set of images.

radius	10	12	14	16	18	20	22	24	26	28
input FFT size 8x8	137	155	163	165	170	163				
innput FFT size 16x16	213	208	216	217	213	215				
input FFT size 32x32	225	229	234	238	237	239				
input FFT size 64x64	15	18	32	49	58	77	80	92	110	109

Figure 4.2: Chart showing the number inliers for different parameters*(inlier is the true matching correspondence).*

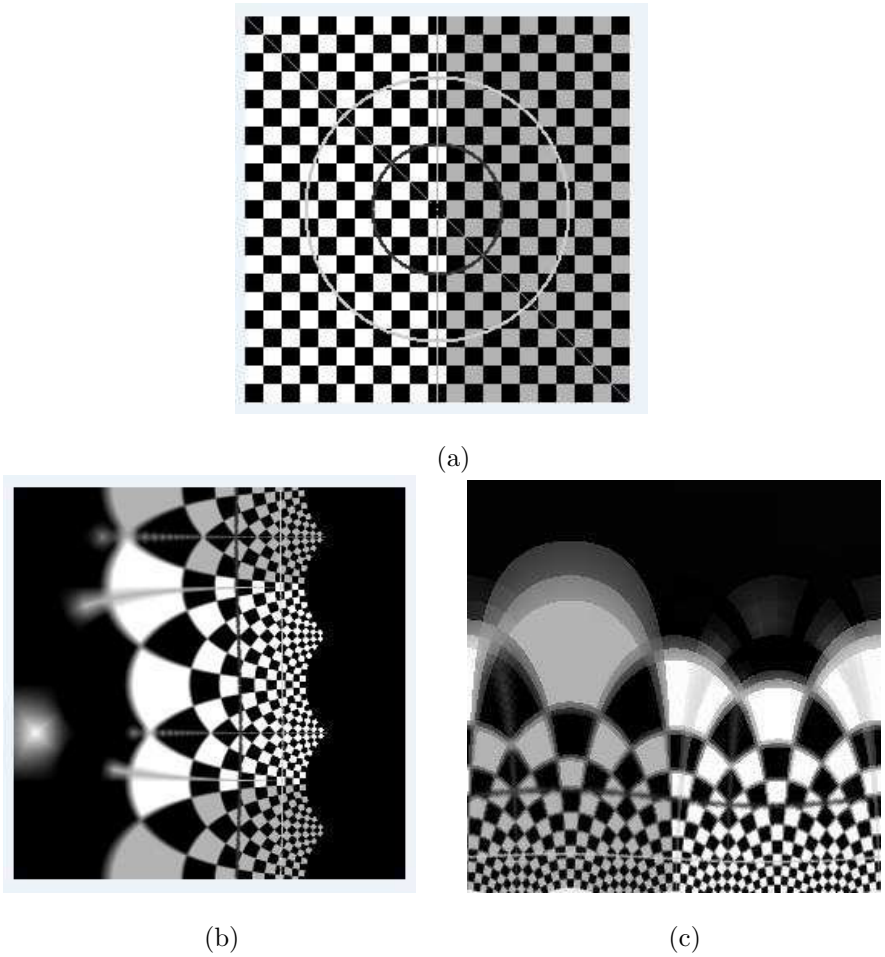
All the tests of the following sections in this chapter are using the optimal parameters— radius for LPT is 20, the input size of FFT is 32x32.

4.2 Results of the LPT in OpenCV and the proposed LPT

In this section the novel log-polar transform and the log-polar transform in OpenCV will be estimated, and their experimental results analyzed. Two sets of samples are given in this section.

Figure 4.3 shows the result after dealing with Log Polar Transform in OpenCV and the proposed LPT. Looking at Figure 4.3 (b), there is a bright spot near the left margin and black semicircles appear on the right border of image. While shown as Figure 4.3 (c), the image appears without a bright spot and black semicircles on the border.

Another experimental result is shown in Figure 4.4. The LPT in OpenCV and the proposed LPT were also tested using Figure 4.4 (a). Looking at Figure 4.4 (b), the back semicircles appear on the border of the image at right side after dealing with LPT in OpenCV. These black semicircles will influence the accuracy of the registration. Figure 4.4 (c) shows now after using the proposed LPT, the image appears without these black semicircles.



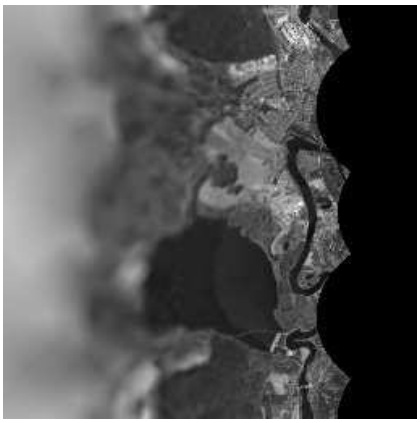
(a) (b) (c)

Figure 4.3: LPT mapping in OpenCV: (a) LPT sampling in Cartesian Coordinates, (b) the sample result using LPT in OpenCV, (c) the sample result using the proposed LPT. (Image source (a) provided by Anders Hast)

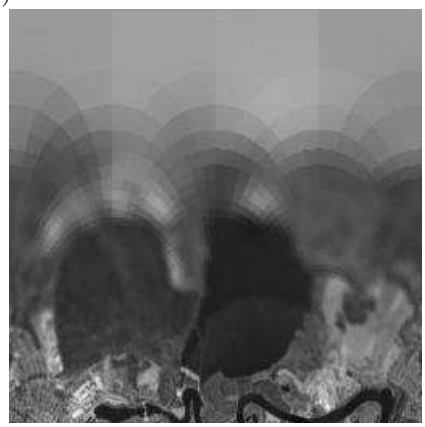
4.3 Results of first log-polar then FFT and first FFT then log-polar

In the experiment, two methods are compared—first log-polar then FFT, and first FFT then log-polar. The novel log-polar method is used to implement the First log-polar then FFT method. Also, log-polar in OpenCV is used to realize the first FFT then log-polar method. Here SIFT is chosen to extract the feature points in both methods. Then the optimal param-

(a)



(b)

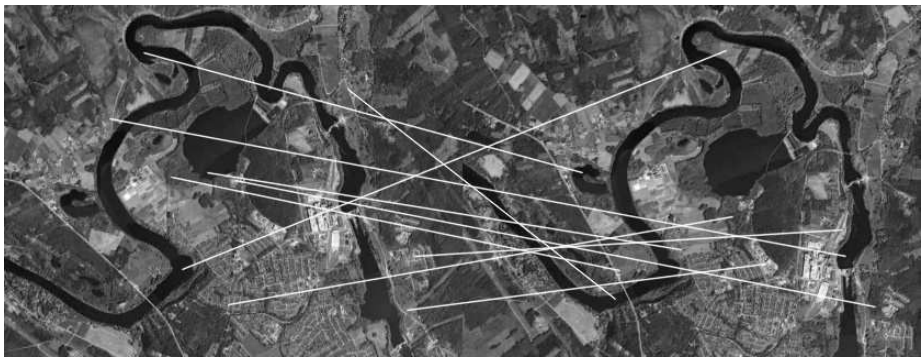


(c)

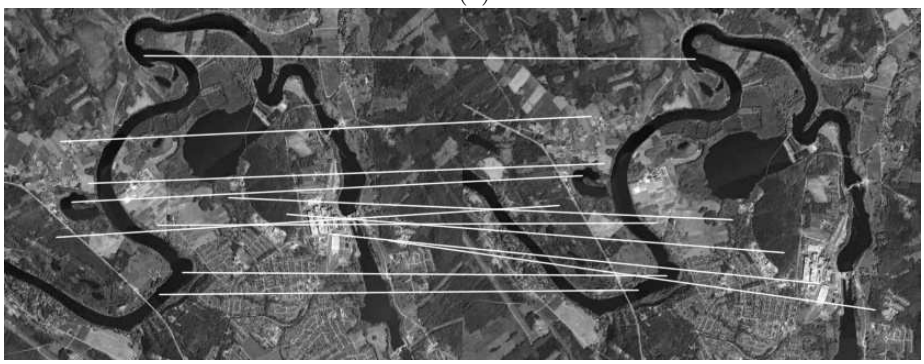
Figure 4.4: LPT in OpenCV and the proposed LPT: (a) LPT sampling in Cartesian Coordinates, (b) the sample result using LPT in OpenCV, (c) the sample result using the proposed LPT.

eters, which are discussed before, are used to complete this experiment.

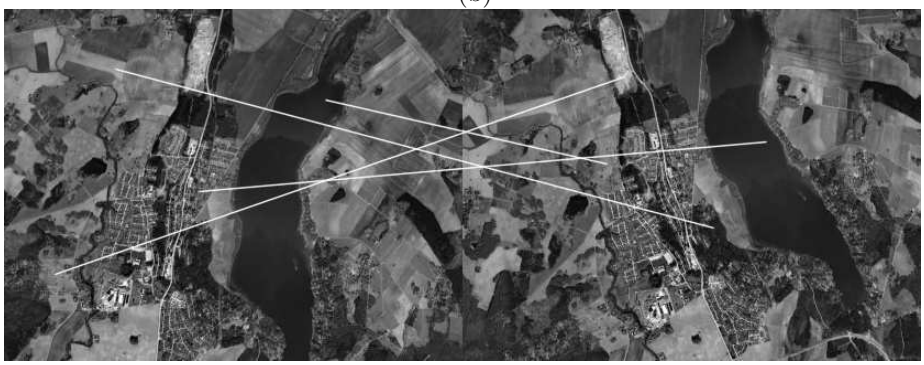
Figure 4.5 shows the results of the two sequence methods—first log-polar then FFT, and first FFT then log-polar. Here the lines are used to connect sets of feature points which have top high phrase correlation. Looking at the matching Connection in Figure 4.5 (a) and (c), there are some false matching connections between the reference image and the sensed image in first FFT then log-polar method. Figure 4.5 (b) and (d) show all the match connections are correct in first log-polar then FFT.



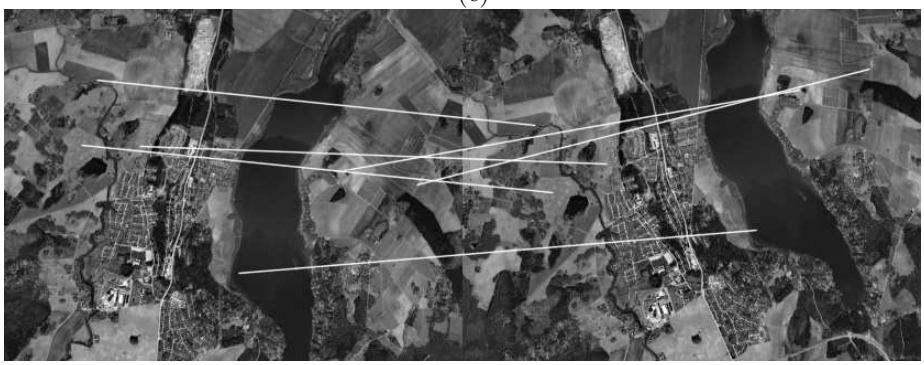
(a)



(b)



(c)



(d)

Figure 4.5: Two sequence method: (a)(c) first FFT then log-polar, (b)(d) first log-polar then FFT.

4.4 Registration Results

The different images are used to evaluate the performance of the novel registration method. The different types of images are described as following:

Normal Image: It is a normal aerial image of resolution 418*320 (133760 pixels). According to this standard image, it has been modified to create other different type of images.

Brightened Image: The normal image have been increased contrast by 20%. The brightened image and the normal image have the same number of pixel.

Rotated Image: The image has been rotated in google earth in different angle degrees– 45 degree, 90 degree, 135 degree, 180 degree, 225 degree, 270 degree and 315 degree. The images have also been rotated in random angle degrees, and then these images have been stored as the sensed image.

Scaled Image: The height and width of the normal image are shrunk by a factor of two using the defaults of bicubic interpolation and antialiasing. The resulting image size is 377*288 (108576 pixels) which contains nearly 90% the pixels of the normal image.

Noisy Image: The normal images, screenshots from Google Earth, are high quality. In order to carry out the test for noise image, salt and pepper noise has been added to the normal image to get the noisy image. Here the noise density 0.02 is chosen.

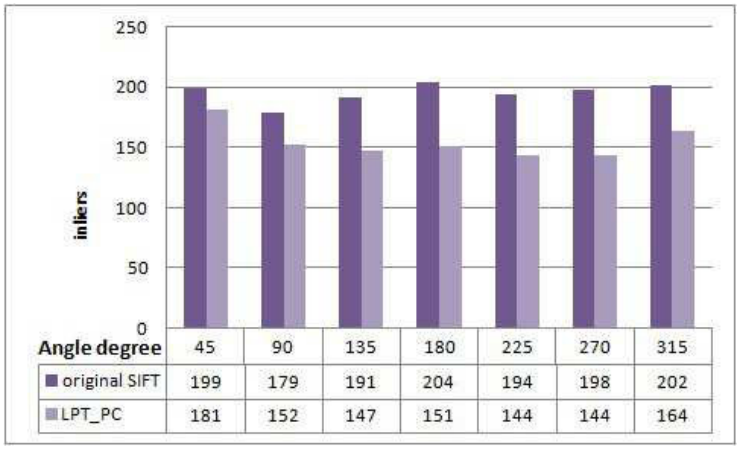
4.4.1 Results of the LPT_PC and the original SIFT

This experiment has estimated the novel registration method LPT_PC with SIFT feature detector. The SIFT is selected to extract the feature points. Next, the first log-polar then FFT method are chosen, which is compared with Lowe's SIFT. Both methods select 400 feature points from the original images. After homography estimation in OpenCV, the number of inliners are compared.

At the beginning, the different angle rotated images are tested in these two registration methods. There are 45 degree, 90 degree, 135 degree, 180 degree, 225 degree, 270 degree and 315 degree rotation images in this experiment. The results of inliers have been presented in the chart shown in Figure 4.8.

As shown in Figure 4.8, the LPT_PC and the original SIFT can cope with rotation images. The original SIFT get the number of inliers above 179, the highest number of inliers is 202. The number of inliers in the LPT_PC are range in 144 to 181.

Here another experimental result is given of the LPT_PC and the original SIFT, which is about random angle rotation image registration. 400 SIFT



Angle degree	45	90	135	180	225	270	315
original SIFT	199	179	191	204	194	198	202
LPT_PC	181	152	147	151	144	144	164

Figure 4.6: Chart showing the number of inliers for different rotation degree*(one image is rotated in different angle degrees).*

feature points are extracted from the reference images and the sensed images. Figure 4.7 shows 4 sets of images with random angle degrees. It also proves the proposed method can handle rotation images.

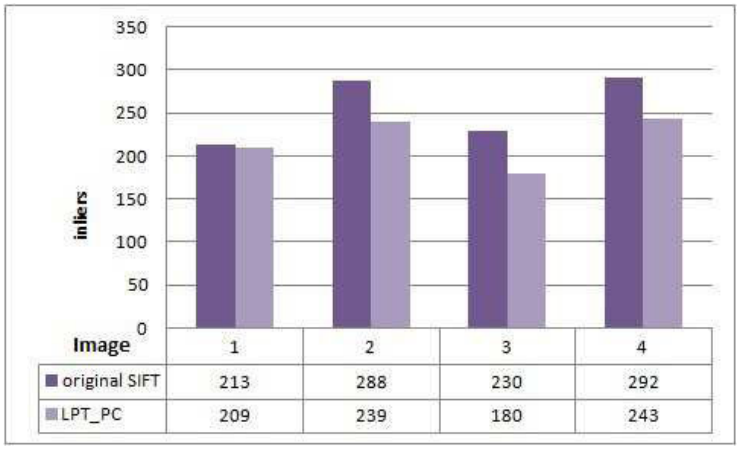


Image	1	2	3	4
original SIFT	213	288	230	292
LPT_PC	209	239	180	243

Figure 4.7: Chart showing the number of inliers for random angle rotation*(Four different images are rotated in random angle degree).*

At last, the number of inliers have been compared in the different type of images — Brightened image, Rotated image (45 degree), Scaled image and Noisy image.

The chart in Figure 4.8 shows that the LPT_PC has high ability for dealing with the Noisy image which got 197 inliers, while the original SIFT gives

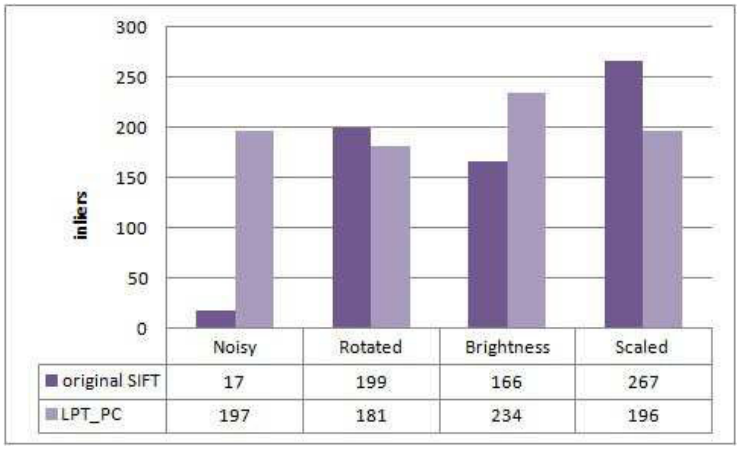


Figure 4.8: Chart showing the number of inliers for different images.

low number of inliers 17. For rotation, both method can cope with it, the inliers in the original SIFT is 199 and in the LPT_PC is 181. For the Brighten image, the inliers in the original SIFT is 166 and in the LPT_PC is 234, meaning The LPT_PC works better. For scale image, the original SIFT performs better than the LPT_PC.

4.4.2 Results of LPT_PC and LPT_CC

Here the novel registration method LPT_PC has been estimated with Harris feature detector. The Harris is selected to extract 400 feature points from the reference image and the sensed image. The first log-polar then FFT method have been chosen. Then two methods have been tested: phase correlation and cross correlation in this case. After homography estimation in OpenCV, the number of inliners are compared.

The chart presented in Figure 4.9 shows the results of inliers of different angle rotation images in these two registration methods. There are 45 degree, 90 degree, 135 degree, 180 degree, 225 degree, 270 degree and 315 degree rotation images in this experiment.

Figure 4.9 shows that LPT_PC obtains more inliers than LPT_PC after rotate large degree.

Another experimental result describes the LPT_PC and LPT_CC dealing with random angle rotation image registration. 400 SIFT feature points are extracted from the reference images and the sensed images. Figure 4.7 shows 4 sets of images with random angle degree. It also proves that the proposed method can handle rotation images. In three sample, LPT_PC works better than LPT_CC. Only one sample shows the LPT_CC is better.

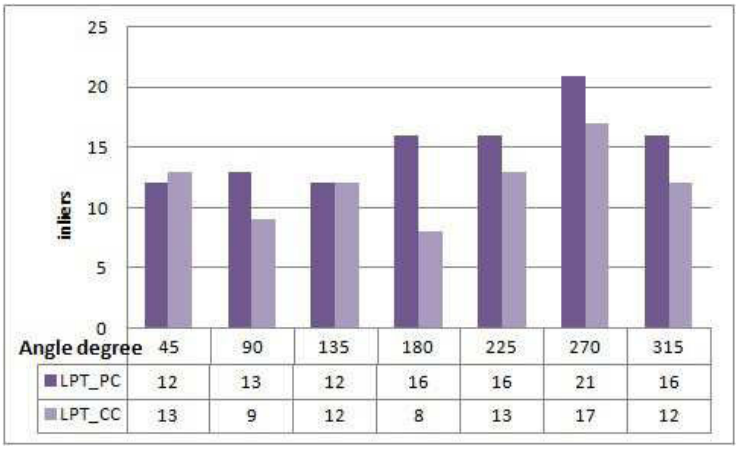


Figure 4.9: Chart showing the number of inliers for different rotation degree using Harris_FFT *(one image is rotated in different angle degrees).*

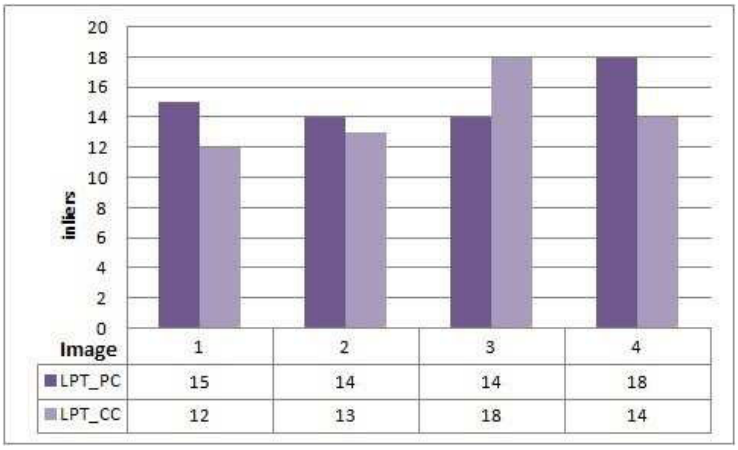


Figure 4.10: Chart showing the number of inliers for random angle rotation using Harris_FFT *(Four different images are rotated in random angle degree).*

Chapter 5

Discussion

A novel Log-Polar Transform (LPT) was used in the proposed registration. Image patches were created using feature points as centers. Each point was a center point of LPT, so circular image patches were cropped over the center points. The circular images were transformed to Log-Polar coordinates. This can avoid black semicircles from appearing on the border of the LPT images.

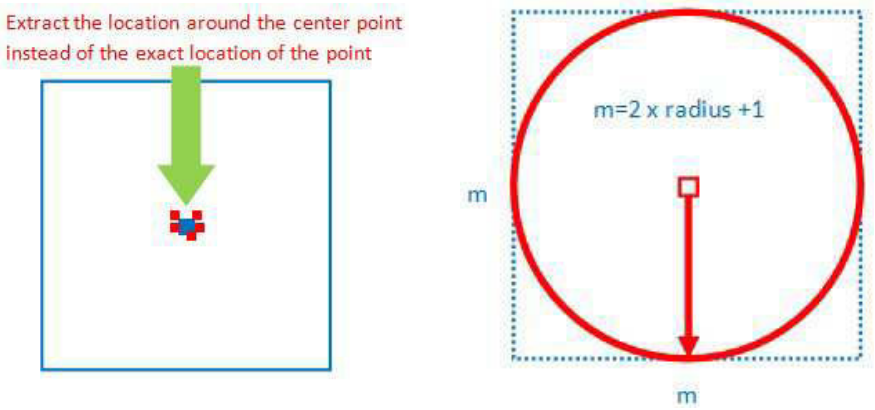
Due to the radius of the circle, image patches can be changed. A lot of image data was tested in the experiments. From experimental results, the best parameters (the radius for LPT as 20, the input size of FFT as 32x32) were found for this novel method. The benchmark was the number of inliers (the true matching correspondence in homography estimation). The best parameters are used in the following registration methods.

In order to evaluate the accuracy of the novel registration. The semicircles and the proposed log-polar were compared at the beginning. The limitation of the log-polar in OpenCV was discovered.

One disadvantage of the log-polar in OpenCV is presented in Figure 4.3. Figure 4.3 shows the result after dealing with Log Polar Transform in OpenCV. Looking at Figure 4.3 (b), there is a bright spot near the left margin. Due to the LPT's principle, the image pixel close to the center is oversampled. If the center of one image is a white bright point in Cartesian coordinates, it will appear as a white bright line on the margin in log-polar coordinate. In this case, it appears as a bright spot. The algorithm in OpenCV extracts the location around the points instead of the exact location of the point. Figure 5.1 shows the principle of the LPT in OpenCV and the proposed LPT. The proposed LPT selects the exact center point. So the LPT in OpenCV is not a ideal solution if a high precision solution is required.

Another disadvantage is described in Figure 4.3 (b) and Figure 4.4 (b), where the black semicircles appear on the border of the image on the

right side after dealing with LPT in OpenCV. These black semicircles will influence the accuracy of the registration. Some solution is required to remove these semicircles.



(a) (b)

Figure 5.1: The principle of the LPT in OpenCV and the proposed LPT: (a) The principle of the LPT in OpenCV , (b) The principle of the proposed LPT

These comparisons indicate that the proposed log-polar has more advantage than the log-polar in OpenCV. As shown in Figure 4.3 (c) and Figure 4.4 (c), the proposed LPT remove these semicircles in success.

Next, the experiments was continued to compare the two methods —first log-polar then FFT, and first FFT then log-polar. The First log-polar then FFT method is implemented by the novel log-polar method. The first FFT then log-polar method is realized by the log-polar in OpenCV.

Here SIFT was chosen to extract the feature points in both methods. Then the optimal parameters (the radius for LPT as 20, the input size of FFT as 32x32) are used to complete this experiment. The results gave a conclusion: first log-polar then FFT is much better than first FFT then log-polar. The novel log-polar method which extract circle image matches is better than the LPT method in OpenCV.

The different images were used to estimate the performance of the novel registration method. From the results of LPT_PC and the original SIFT, both methods can cope with rotated images. Although the original SIFT can get more inliers, the LPT_PC still deals well with the different angle rotated images. For Noisy images, the LPT_PC surpasses the original SIFT a lot. For Brightened images, the LPT_PC still performs better than the original SIFT. The reason is the brightness change was described

as the phase difference in frequency domain, phase correlation can cope with the transforming shift as phase difference in the Fourier domain. This experiment tested the normal image increased contrast by 20%. If more increasing contrast appear, it requires bigger radius of the circle in image patch to operate in phase correlation step. For a scaled image, the original SIFT works better. The LPT_PC still works on it when the scaling factor is small (the scaling factor as 0.8 in this experiment). The LPT_PC will fail if the scaled image contains nearly half the pixels of the normal image.

At the end of experiments LPT_PC and LPT_CC were discussed. Both methods are tested in different angle rotated images. It indicated that the novel method can handle rotated images. The LPT_PC works better than LPT_CC.

For the high accuracy requirement our registration method suffered from computational complexity: Harris is faster at detection than SIFT. For removing the false matches, the matches are computed with a symmetry test. After the symmetry test, SIFT got more accurate match corner than Harris.

The homography matrix is a suitable transformation model for 2D aerial images. RANSAC computes a homography requiring at least 4 correspondences. Here RANSAC was chosen instead of Least Median of Squares Regression (LMS), since LMS does not deal with the outliers, which is more than half the data. It is important that the resulting transformation is free of outliers, so that it will not distort the image.

Sub-pixel precision is tried in the novel registration method, the results being worse than integer pixel precision. Therefore, The novel method chose integer pixel precision to complete registration.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This thesis contains an overview of image registration methods and proposes a novel registration method which can be used for 2D aerial images. The aim of the work was to improve the accuracy of registration. It selected the feature points with corner detector and created image patches using feature points as centers. Instead of using the whole image in the conventional method, feature points are used in this project, which reduces the computational time. For rotation invariant, it is not important how the image patch is rotated. The key is focusing on the feature points. Each point was a center point of LPT, so the circular image patches were cropped over the center point. Then it transformed the circular images to Log-Polar coordinates, which can avoid black semicircles from appearing on the border of the LPT image. Next, it coped with the LPT images with phase correlation.

Due to the radius of the circle, image patches can be changed. After testing, the best parameters have been found (the radius for LPT as 20, the input size of FFT as 32x32) for this novel method.

Image registration with FFT always do the FFT first. But in our case it is bad as we do not want the corners and not the black semicircles either. Therefore, I do LPT first to get the circular region and then the FFT. Comparison of two methods — first Log-polar then FFT or first FFT then Log-polar. Furthermore, the novel LPT with the LPT in OpenCV have also been compared. Two disadvantage of the LPT in OpenCV were discovered.

This novel method (novel Log-Polar Transform and first Log-polar then

FFT) have been compared with Lowe's SIFT in different types of images with noise, rotation, high intensity and scaling. The results were analyzed and evaluated from this novel registration method. The experiment also tested the LPT_PC (Log-Polar Transform and Phase Correlation), and LPT_CC (Log-Polar Transform and Cross Correlation). According to the number of inliers, the LPT_PC works better.

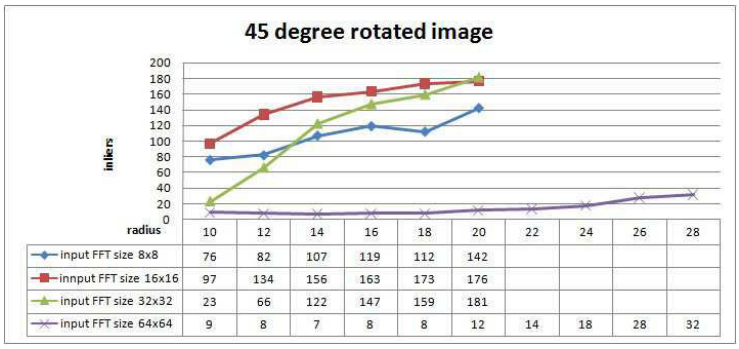
In one word, I have implemented a reliable novel method for 2D aerial images registration. Comparing with the previous common method, this novel registration method is able to solve the 2D aerial images in different rotated images at the present stage. It copes well with the images in different locations at the same time and finds the transformation between the reference image and the sensed image. Especially in Noise images, its performance surpasses the Lowe's SIFT a lot.

6.2 Future Work

To build more reliable and robust registration, the extension of this project will be focused on obtaining the optimal parameters for old aerial photos from different time period. It is important to research the historic buildings change. For example, given several images in different time — 60's, 70's, 80's and 90's, the registration method could find the same objects and tell about the historic changing among the given images. This method could be applied in tourism, where the development of a place of interest could be followed. People can easily understand the history in the place of interest.

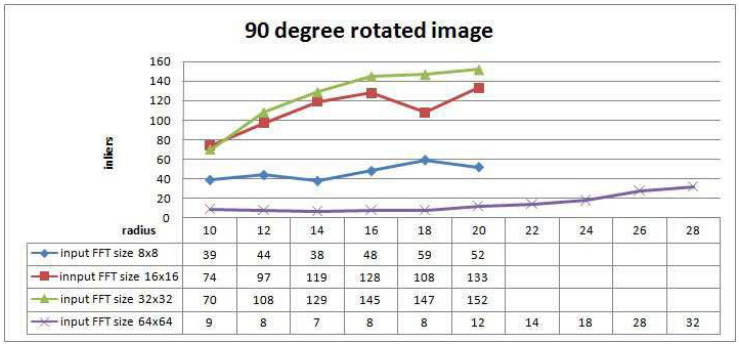
Furthermore, the computational complexity could be reduced and the registration made faster and more efficient. Python is a interpreting programming language, which incorporates easy with other application. However, Python is too much slower than C++. The registration method could be developed by C++ if fewer running time is required. The accuracy of the novel registration needs to be improved as well.

Appendix A



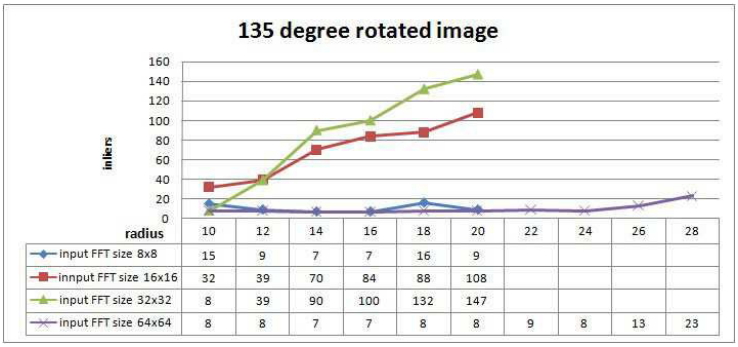
radius	10	12	14	16	18	20	22	24	26	28
input FFT size 8x8	76	82	107	119	112	142				
innput FFT size 16x16	97	134	156	163	173	176				
input FFT size 32x32	23	66	122	147	159	181				
input FFT size 64x64	9	8	7	8	8	12	14	18	28	32

Figure A.1: Parameters testing for 45 degree rotated image



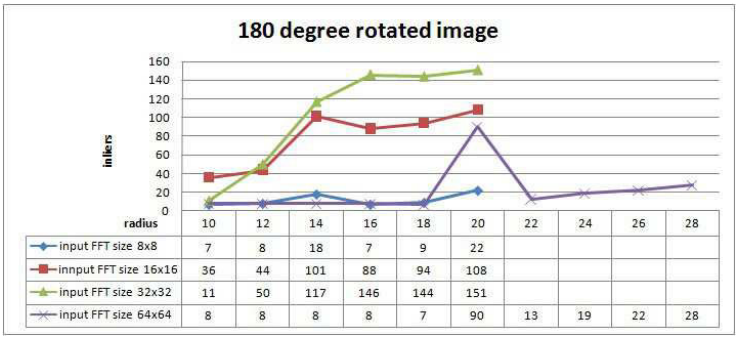
radius	10	12	14	16	18	20	22	24	26	28
input FFT size 8x8	39	44	38	48	59	52				
innput FFT size 16x16	74	97	119	128	108	133				
input FFT size 32x32	70	108	129	145	147	152				
input FFT size 64x64	9	8	7	8	8	12	14	18	28	32

Figure A.2: Parameters testing for 90 degree rotated image



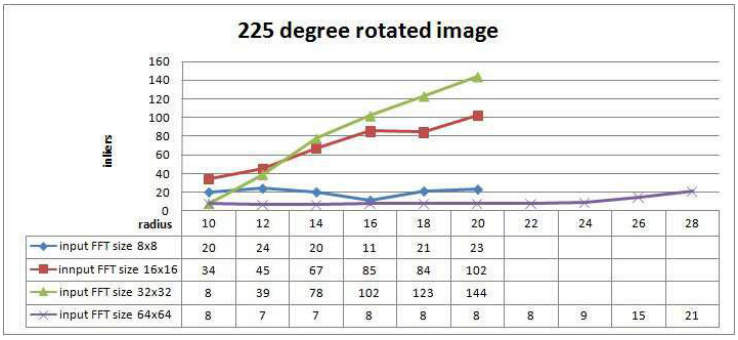
radius	10	12	14	16	18	20	22	24	26	28
input FFT size 8x8	15	9	7	7	16	9				
innput FFT size 16x16	32	39	70	84	88	108				
input FFT size 32x32	8	39	90	100	132	147				
input FFT size 64x64	8	8	7	7	8	8	9	8	13	23

Figure A.3: Parameters testing for 135 degree rotated image



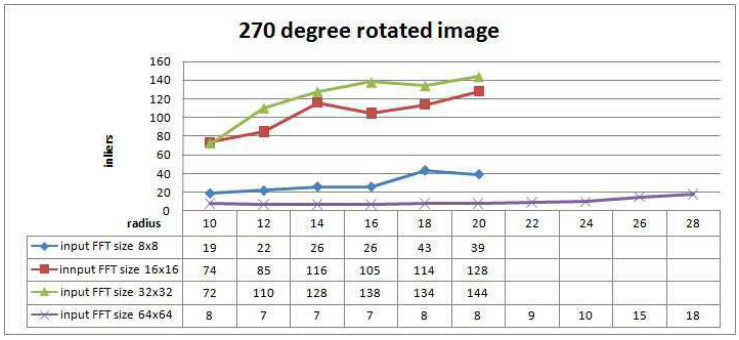
radius	10	12	14	16	18	20	22	24	26	28
input FFT size 8x8	7	8	18	7	9	22				
innput FFT size 16x16	36	44	101	88	94	108				
input FFT size 32x32	11	50	117	146	144	151				
input FFT size 64x64	8	8	8	8	7	90	13	19	22	28

Figure A.4: Parameters testing for 180 degree rotated image



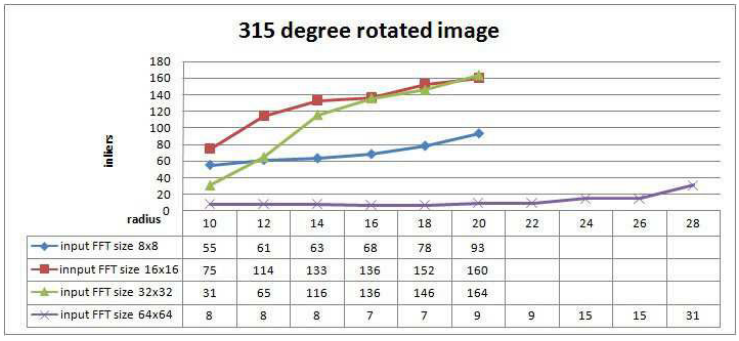
radius	10	12	14	16	18	20	22	24	26	28
input FFT size 8x8	20	24	20	11	21	23				
innput FFT size 16x16	34	45	67	85	84	102				
input FFT size 32x32	8	39	78	102	123	144				
input FFT size 64x64	8	7	7	8	8	8	8	9	15	21

Figure A.5: Parameters testing for 225 degree rotated image



radius	10	12	14	16	18	20	22	24	26	28
input FFT size 8x8	19	22	26	26	43	39				
innput FFT size 16x16	74	85	116	105	114	128				
input FFT size 32x32	72	110	128	138	134	144				
input FFT size 64x64	8	7	7	7	8	8	9	10	15	18

Figure A.6: Parameters testing for 270 degree rotated image



radius	10	12	14	16	18	20	22	24	26	28
input FFT size 8x8	55	61	63	68	78	93				
innput FFT size 16x16	75	114	133	136	152	160				
input FFT size 32x32	31	65	116	136	146	164				
input FFT size 64x64	8	8	8	7	7	9	9	15	15	31

Figure A.7: Parameters testing for 315 degree rotated image

References

[1] P.E. Anuta. Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques. *Geoscience Electronics, IEEE Transactions on*, 8(4):353–368, Oct.

[2] H. Araujo and J.M. Dias. An introduction to the log-polar mapping [image sampling]. In *Cybernetic Vision, 1996. Proceedings., Second Workshop on*, pages 139 –144, dec 1996.

[3] Daniel I. Barnea and H.F. Silverman. A class of algorithms for fast digital image registration. *Computers, IEEE Transactions on*, C-21(2):179–186, Feb.

[4] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. 3951:404–417, 2006.

[5] Ronald N. Bracewell. *The Fourier Transform and Its Applications, 2nd ed.* McGraw-Hill, USA, 1986.

[6] C.D.Kuglin and D.C.Hines. The phase correlation image alignment method. *proc.Cybernetics and Society*, pages 163–165, September 1975.

[7] Qin-Sheng Chen, M. Defrise, and F. Deconinck. Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(12):1156 –1168, dec 1994.

[8] Elan Dubrofsky. Homography estimation. Master's thesis, University of British Columbia, 2009.

[9] Danijela Ristic-Durrant Faraj Alhwarin, Chao Wang and Axel Grser. Improved sift-features matching for object recognition. *BCS International Academic Conference 2008 Visions of Computer Science*, pages 179–190, 2008.

[10] J.B. Zerubia Foroosh, H. and M. Berthod. Extension of phase correlation to subpixel registration. *IEEE Transactions of Image Process*, Volume 11, Mars 2002.

[11] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.

[12] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[13] Anders Hast, Andrea Marchetti, Beatrice Rapisarda, J Sheperd, and Maurizio Tesconi. Geomemories - a spatial-temporal atlas of the italian landscape. In *International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, pages 41–44, 2011.

[14] Luo Juan and Oubong Gwon. A Comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing (IJIP)*, 3(4):143–152, 2009.

[15] F. Jurie. A new log-polar mapping for space variant imaging.: Application to face detection and tracking. *Pattern Recognition*, 32(5):865 – 875, 1999.

[16] Yan Ke and R. Sukthankar. Pca-sift: a more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–506 – II–513 Vol.2, june-2 july 2004.

[17] David G. Lowe. Object recognition from local scale-invariant features. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, Corfu, Greece, 1999.

[18] David G. Lowe. Local feature view clustering for 3d object recognition. In *Proceedings of the 2001 IEEE Computer Society Conference on omputer Vision and Pattern Recognition*, Kauai, Hawaii, 2001.

[19] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[20] R. Matungka, Y.F. Zheng, and R.L. Ewing. 2d invariant object recognition using log-polar transform. In *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*, pages 223 –228, june 2008.

[21] R. Matungka, Y.F. Zheng, and R.L. Ewing. Image registration using adaptive polar transform. *Image Processing, IEEE Transactions on*, 18(10):2340–2354, 2009.

[22] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10):1615–1630, 2005.

[23] Sankar K Pal Nikhil R Pal. A review on image segmentation techniques. *Pattern Recognition*, Volume 26:1277–1294, September 1993.

[24] P.J.Rousseeuw. Least median of square regression. *Journal of the American Statistical Association*, 79(388):871–880, 1984.

[25] Chi-Man Pun and Moon-Chuen Lee. Log-polar wavelet energy signatures for rotation and scale invariant texture classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5):590 – 603, may 2003.

[26] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Computer Vision, 1998. Sixth International Conference on*, pages 230–235, Jan.

[27] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Evaluation of interest point detectors. *Int. J. Comput. Vision*, 37(2):151–172, June 2000.

[28] Utkarsh Sinha. Sift: Scale invariant feature transform. http://www.aishack.in/2010/05/sift-scale-invariant-feature-transform, May 2010. Date visited: 2013-01-20.

[29] S. M. Smith and J. M. Brady. Susan - a new approach to low level image processing. *International Journal of Computer Vision*, 23:45–78, 1995.

[30] Tinne Tuytelaars and Krystian Mikolajczyk. *Local Invariant Feature Detectors: A Survey*. Now Publishers Inc., Hanover, MA, USA, 2008.

[31] Hongshi Yan and Jian Guo Liu. Robust phase correlation based feature matching for image co-registration and dem generation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII. Part B7, 2008.

[32] Fei Yuan, Hong Zhang, and Ruiming Jia. Digital image stabilization based on log-polar transform. In *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*, pages 769 –773, aug. 2007.

[33] Barbara Zitov and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, October 2003.

[34] S. Zokai and G. Wolberg. Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations. *Image Processing, IEEE Transactions on*, 14(10):1422 –1434, oct. 2005.