SHAPE CONTEXT DESCRIPTOR: A NEW APPROACH TO ASSESS COSISTENCY OF TWO CORRELATED LAND BOUNDARIES

by

Chong Chi Man

Supervisor: Dr. YAN Wai-Yeung

Moderator: Dr. Liu Xin-Tao

A dissertation

submitted to The Hong Kong Polytechnic University

in partial fulfilment of the requirement

for the degree of

Bachelor of Science (Honours)

In the Program of

Geomatics (Land Surveying)

May 2020

Certificate of Originality

I hereby, declare that this dissertation is my own work and that, to the best of my knowledge and belief, it contains no material which has been accepted or submitted for the award of any other degree or diploma, except where due reference is made in the text of the text.

Abstract

This dissertation presents a new approach for the correlating land boundaries using shape contexts descriptor. In Hong Kong, the current practise for the boundary correlation is conducted through human interpretation, which is more likely a subjective approach. Therefore, I explore the use of shape context to correlate the boundaries in land records to assist land surveyors in boundary survey work. The boundary of the same lot in two land records can be identified and matched based on its geometry. A coordinate transformation model can be established using the matched vertices of the two boundaries. Least squares adjustment is used to estimate the model coefficients. Accuracy assessment of the result can be performed by Dice Similarity Coefficient (DSC). The proposed method was tested on the lots in Demarcation District (DD) Sheet 1 TC, 129 and the corresponding Survey Record Plans (SRP).

Acknowledgment

I would like to express my sincere gratitude to Dr. Yan Wai-yeung in providing me with a suggestion on applying shape context descriptor into Hong Kong Cadastral Survey System. I remembered the first time I meet with him was around July 2019. Since I was going to join an exchange study in Calgary from September 2019 – December 2019, I would like to come up with a topic earlier. I told him I am interested in topics related to land boundary survey. I assumed that I would have my dissertation on a court case study related to the land boundary dispute, or research on surveying techniques. These are the usual topics, which done by students in previous years. Surprisingly, Dr. Yan suggested me to do research on computer vision algorithm and apply it into current land boundary survey system. This was what I never think of. With his kind assistance, I came up with an idea to apply shape context descriptor in the correlation of boundaries. Then I started my studies in this area. During my exchange study, I spent a lot of time on the study at the University of Calgary. The progress of my dissertation was a little bit slowed down.

As soon as I came back to Hong Kong, I continued to do my dissertation. Face-to-face teaching was suspended due to the coronavirus. The work became challenging since I could not meet Dr. Yan face-to-face. It was not efficient to communicate through email. I still manage to finish the dissertation because of him. He called me and even held a video meeting to assist me when I faced any difficulties. Also, I would like to appreciate Dr. Matthew Pang. He is the one who managing the overall writing of the dissertation. He taught me about general writing skill and advised me some of key issues on my dissertation. Furthermore, I would like to appreciate the assistance form Dr. Eric Tang and Ms. Stella Tse. Dr. Tang is an authorised land surveyor. He invited me to go to his office when he knew that I might need his help. Ms Tse is his colleague. She helped me to find the land records of the 40 lots. Without these data, I may need to spend a lot of time and a thousand of money for searching the documents. Moreover, I would like to thank my family. I appreciate Ms. Sun Wun Ki for her patience. She encouraged me when I felt stressful in doing the dissertation. I will not finish this dissertation without the supports from them.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Research Motivation

On 9th June 1898, an agreement which was known as "Convention for the Extension of Hong Kong Territory", was signed between the Qing dynasty of China and the United Kingdom. Under the agreement, Great Britain was granted an additional 99 years of rule over the Hong Kong colony. North of Boundary Street, south of the Sham Chun River and the surrounding islands, with a total area of 376 square miles, were leased to the United Kingdom for 99 years (Couling, 1917). The area is known as New Territories later. After the British took over Hong Kong, the Colonial Secretary of Hong Kong, Sir Janes Haldane Stewart Lockhart conducted an all-round investigation, regarding the development of the colony of Hong Kong. In his report, he emphasized that the first issue needed to tackle was to settle the land title. His view was then agreed by the Secretary of State for the Colonies, Sir Joseph Chamberlain (Colonial Office, 1900). Chamberlain assigned Lockhart to handle the work of cadastral survey. This was the starting point of the Demarcation District (DD) survey.

To solve the problem of land title and rent collection, land registration system was introduced in New Territories. To establish the land registration system, DD survey was then conducted from 1899 to 1904. The survey team was led by two land surveyors: Mr. G.P. Tate and Mr. W.J. Newland, together with the survey officers from Indian. The maps created in DD survey were known as "DD Sheets". For the colonial government, DD Sheets were adopted to settle the land title and collect rent. They were adequate to serve the need at that time but not in today's city development. Due to the increase of land value, the boundary determination of the lots became an important issue. Re-establishment of lot boundary can help people to identify the exact location of the land so they can do further development or tickle with the boundary dispute problem.

Re-establishment of lot boundary is not an easy task, which involves several steps. Land records, aerial photograph, current survey results are the information needed to determine a boundary. One of the procedures is to check if the surveyed area agrees with the DD Sheet's area, and the registered area to within a certain tolerance. However, there is no such criteria to assess the shape and orientation between the boundaries of the surveyed area and the area depicted in DD sheet. The area between the two boundaries can be the same but their shapes can be different. Therefore, this dissertation introduces a computer vision algorithm, shape context descriptor, to determine the similarity of the two boundaries. Dice Similarity Coefficients (DSC) can be used for accuracy assessment. Two DD Sheets plotted in different scale are chosen in this experiment. This aim to have a comparison of the shape context's performance in the plans plotted in a different scale.

1.2 Research Objectives

The ultimate goal of this research aims to develop an algorithm to perform the correlation between two lot boundaries. In the current practice, the whole correlation processes are done by a human. Land surveyors will digitise the boundary in the first plan and second plan, respectively. The two digitised boundaries can be put together to perform the correlation. They will make an interpretation of the two boundaries' shape according to the correlation result: are they consistent with each other? All these processes may contain human error and the interpretation is a subjective statement. In shape context descriptor, the two digitised boundaries will be correlated by matching vertices and applying transformation algorithm, which can reduce human error. A value known as DSC representing the similarity between two shapes can be computed. Land surveyors can use it as a reference to make the interpretation of the correlation results. Specifically, the objectives of the research are:-

1. To use the shape context descriptor to do the correlation of lot boundaries so as to reduce the human error;

2. To compare the correlation result by using shape context descriptor for different scale of DD sheet; and

3. To provide a new method for land surveying industry regarding the correlation of the lot boundaries, in an objective way.

1.3 Organization of the dissertation

This dissertation is organized as follows: Chapter 2 presents the literature of the procedure of the boundary re-establishment, and prior work on shape context matching. Chapter 3 presents the methodology of using shape context descriptor on two shapes. Sample data, size and purpose of the experiment are presented. Chapter 4 presents the experimental result to assess the feasibility of using shape context on the correlation of land boundaries. Chapter 5 presents the conclusion of the research and the directions for future work.

Chapter 2

Literature Review

2.1 Defects of Land and Survey Records

There are different land and survey records to be used during the boundary re-establishment. They can be classified into 3 types of evidence: primary evidence, secondary evidence and extrinsic evidence. Lands in Hong Kong are mainly granted in two ways: Old Schedule Lot and New Grant Lot. The evidence used to establish these two types of the lot is different, especially the primary evidence. The primary evidence for the boundary re-establishment of Old Schedule Lots is the Demarcation District (DD) Sheet attached to Block Government Lease. It has a higher priority over other evidence. It shows the topological relationship of the lot has depicted on the sheet, the number of cultivation fields is included in the lot, the adjoining relationship. For the New Grant Lot, the primary evidence used for the boundary re-establishment is the New Grant Register and Lease Plan. However, all the primary evidence has defect to show the exact location of the subject lot, especially the primary evidence of Old Schedule Lots. Defects of DD sheet will be explained in this section.

DD sheet is the primary evidence of Old Schedule Lot, which was a product from DD survey. To explain the defect of DD Sheet, DD survey should be introduced first. DD survey started in 1899 and ended in 1904. It was divided into 2 parts: triangular survey and cadastral survey. The triangular survey was conducted from 1899 to 1900. 10 trigonometric stations were established and used by the cadastral survey team. The cadastral survey was further divided into 3 parts: traverse survey, detail survey and demarcation survey.

The purpose of the traverse survey was to establish the traverse stations for the detail survey team to use. Mr. Newland have some comments about the survey instruments used during the traverse survey: -

> "Five-inch theodolites of the Everest pattern were used for the angular work; and for the measurements of all distance, we used 2 chains of unequal lengths, one being the ordinary Gunters chain of 100 links = 22 yards, and other a Gunters chain minus 7 links" (Newland, 1904).

It is observed that primitive survey methods were being used. However, it does not mean that the accuracy of the traverse survey is low. He further implied that the traverse survey result was verified and checked: -"To ensure the accuracy of all chaining, each Surveyors squad, in addition to the ordinary working chains, was supplied with 2 standard chains. All chains before being made over to the Surveyors, were tested on the standard laid down at the Headquarter camp, by means of the steel standard tape, and the spare chains were sent in once a fortnight to be tested on the standard. The surveyors had orders to test their working chains daily in the field against their spare standard chains" (Newland, 1904).

Traverse survey was finished in 1902. A totally 8928 of traverse station, 87468 chains or 1093.35 meters were established. Mr. Newland also expressed his feelings towards the result of the traverse survey. He felt fortunate that he was able to keep the traverse survey ahead of the cadastral survey. This implied that DD survey was conducted in a hurry, which may increase the chance of creating error. In conclusion, the result of the traverse survey was satisfactory. The traverse survey was well and successfully conducted, respecting to its accuracy requirement and the purpose.

Detail survey is the key factor for the defect of DD Sheets. In detail survey, 328639 lots in 477 DD were surveyed, with a total area exceeding 40000 acres. The maps were produced in two scales: 471 of maps were plotted at 1:1980 while 85 of maps were plotted at 1:3960 (Newland, 1904). These maps are known as "DD Field Sheets". Leung (2008) commented on DD Sheets: "The DD Sheets were not too bad for achieving their function one hundred years ago." True, DD Sheets in such a small scale were good enough for the registration and rent collection. For map plotted at 1:1980, 1 mm on the map represents 2 m on the ground. It was even worsened on the map in 1:3960, which 1mm on the map represents 4 m on the ground. The uncertainty is very high, not to mention the drawing error induced by the tracers. It was estimated that two major factors affected the quality of DD survey, thus the quality of DD Sheets was also affected. The two factors are the survey techniques and time.

Plane-tabling method was used in the detail survey. There are both advantage and disadvantage for the plane-tabling method. For the advantage, less time was spent by using the plane-tabling method. Map of the area is plotted in the field, which means that the fieldwork and the plotting are done simultaneously. Also, the plane-tabling method can be done in less labour, which means more groups can conduct the survey at the same time. At that time, the problem of land title in New Territories has to be solved as soon as possible. The plane-tabling method fit the requirement from the British Colony. Over 40000 acres of land were surveyed within 5 years (Newland, 1904).

For the disadvantage, plane-tabling method produces a lot of error, including instrumental errors, error in plotting, errors of manipulation and sighting (Anupoju, 2019). It is not very accurate for a large-scale survey, comparing to the theodolite survey. It is difficult to create a large-scale map by using this primitive survey method. Other than the defects of the survey method, the performance of the survey officers was not satisfactory during the DD survey. There were several reasons: challenging survey area, poor hygienic situation and inadequate supply of trained staff. Mr. Tate explained this problem in the survey report: -

"The nature of the country is broken and mountainous, and the greater part is some of the most difficult country that I have ever seen. Indian experience does not help one very much, as the conditions are so very different in the New Territory" (Hong Kong Government, 1902d, p.708).

2.2 Lots Boundary Re-establishment

Nowadays, boundary re-establishment of Old Schedule Lots and New Grant Lots is one of the major tasks for land surveyors in Hong Kong. It is a method in tackling the historical problem created by DD survey. DD sheet no longer served the need for today's development in Hong Kong. There are few reasons for raising the demand of lot boundary re-establishment work. One of the reasons is the raising of land value. People can realise the extent of his or her land after land boundary survey and expel the trespassers. The other reason is about land development. In Hong Kong, there are three controls monitoring land development: land covenant, building control and planning control. If all these restrictions have been solved, a General Building Plan (GBR) is required and submitted to the Buildings Department. A land survey plan stating the lot area and coordinates of the boundaries was required to be attached with GBR.

It is better to briefly explain the land registration system first, before introducing the detailed procedure of lot boundary re-establishment. The current practising land registration system is the deed registration system. It was implemented by Colonial Government after DD survey, in 1905 They requested the lot/parcel owners to bring the White Deeds or Red Deeds to Land Court. Land court will then process the deed registration and give the land title to the villagers. The Land Court will also deal with the dispute claims from the villagers. After writing down the lot ownership, and necessary revision, DD Field Sheets were traced and formed as DD Sheets. Then, the parcels in the New Territories were officially granted under the Block Government Lease. These lots were known as Old Schedule Lots. The Block Government Lease is composed of Indenture, Schedule of Government Lessees, and DD sheet. The detail of the DD sheet was explained in Section 2.1. For the Indenture of each BGL, it states that a piece of land parcel was "more particularly delineated and described on the plan attached here to according to the lot number set out in the Schedule" (Lands Department, 2019). DD sheet is the "plan" stated in the above sentence.

A boundary re-establishment can be divided into three main parts: searching of land records, field survey, correlation. In the correlation part, the re-established boundary will be correlated onto different survey records and aerial photo. The purpose is to check the consistency between the surveyed area and the other survey record or historical features. In section III of Land Survey Ordinance, five general principles for re-establishment of lot boundaries was stated. The first principle is to follow the intention of grant: -

> "The re-established boundaries (in terms of position and dimension) and the resultant shape of the lot should not deviate materially from that as shown on the land grant document (Lands Department, 2006)."

There are three key points in the above statements: position, dimension and shape.

Dimension and shape can be classified as the same group, since shape changes as dimension changes. The dimension of a boundary can be defined as length and width, which can form an area. Therefore, area can become a reference to determine the consistency of the dimension and shape of the boundaries. To ensure the re-established boundary is consistent with the boundary stated in the land grant document (i.e. DD sheet), the Lands Department established a standard to check the area discrepancy: -

For DD sheets of 1:1980, the discrepancy is suggested to be within:

+/- (20 sq. metres + 5% of the registered area)

For DD sheets of 1:3960, the discrepancy is suggested to be within:

+/- (20 sq. metres + 10% of the registered area)

However, the standard has a defect. Assumed that there are two boundaries: boundary A and boundary B. The have the same area of 1000 sq. m. By the above standard, there is no discrepancy. However, boundary A has a length of 50 m and a width of 20 m, while boundary B has a length of 100 m and a width of 10 m. Under this situation, land surveyors will give the interpretation for the consistency based on their professional experience and other boundary evidence. Unlike the area aspect, there is no standard showing the shape discrepancy. Therefore, shape context descriptor is introduced to determine the similarity between two boundaries. DSC can act as a reference to assist the land surveyors on their judgement. Figure 2-1 shows a part of workflow of boundary re-establishment of Old Schedule Lot. The boxes presented the procedure in the current practise while the grey boxes presented the new implemented approaches based on the shape of the boundary.

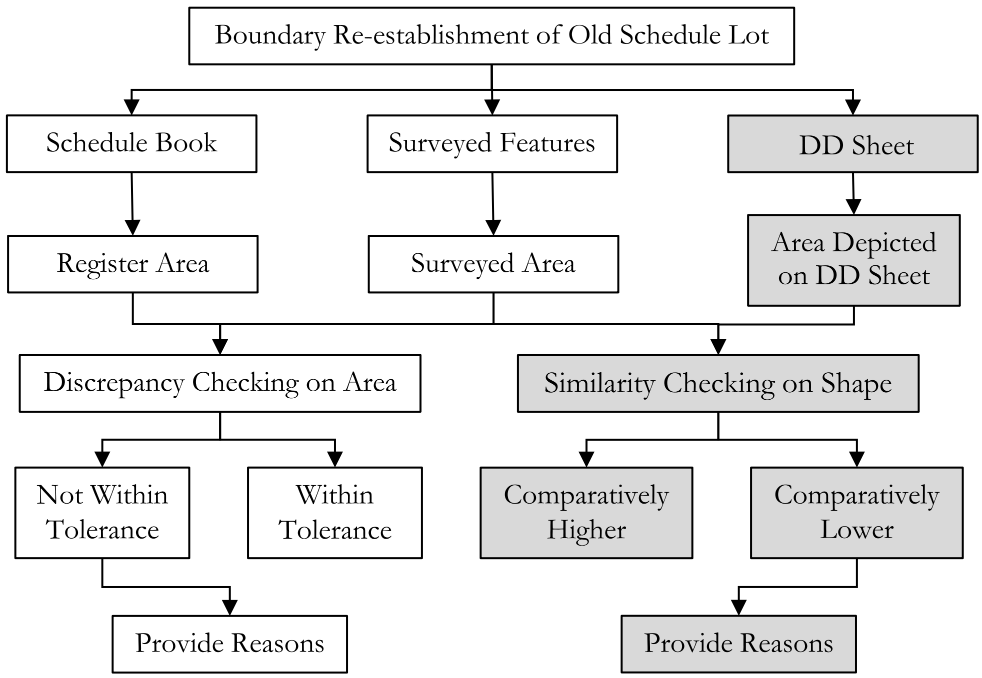


Figure 2-1: Discrepancy checking on area (white boxes) and similarity checking on shape (grey boxes)

2.3 Shape Context and Dice Similarity Coefficient

Shape context is a descriptor used in object recognition, which can be used for measuring shape similarity and recovering point correspondences. In these years, there is an increasing number of studies talking about the theory and application of shape context. Belongie and Malik (2000) were the earliest persons to propose this computer vision algorithm. Shape context is used to describe the coarse arrangement of the shape with respect to the points inside or on the boundary of the shape. The algorithm then combined with a conventional appearance-based descriptor such as the local orientation. The combined algorithm was operated on a 2D image. Correspondences and a similarity score were produced and used for object recognition and similarity-based querying. Belongie, Mori and Malik (2006) further illustrated the use of shape context algorithm on object recognition. The used two handwritten digits as an example (Figure 2-2). By human judgement, these two images seem to be similar. However, they are quite different, in terms of pixel-to-pixel comparisons. Therefore, shape context can be used to determine the shape's similarity.

Figure 2-2: Examples of two handwritten digits. Adapted from Belongie, S., Mori, G., Malik, J., Krim, H., & Yezzi, A. (2006). Matching with Shape Contexts. In Statistics and Analysis of Shapes (Modeling and Simulation in Science, Engineering and Technology, pp. 81-105). Boston, MA: Birkhäuser Boston.

The measurement of shape's similarity can be preceded in three stages (Belongie, Mori and Malik, 2006). The first stage is to solve for correspondences between points on the two shapes. Shape context descriptor will be attached to each and capture the distribution of the remaining points relative to a reference point, thus offering a globally discriminative characterization. If two shapes are similar, their corresponding points will have similar shape contexts. The second stage is to use the correspondences to estimate an aligning transform. A transformation that best aligns the two shapes will be estimated. This can be achieved by using regularized thin-plate splines, which provides a flexible class of transformation maps. The last stage is to calculate the distance between two shapes as a sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transform.

Shape context can also be applied for maps geo-referencing. Yan, Easa and Shaker (2017) proposed a polygon-based image registration method to georeference historical maps. It is obvious that point features are commonly used as control primitives to estimate model parameters in remote sensing and Global Position System. Some researches show the use of linear and curvy features for remote sensing image registration. For geo-referencing of historical map data, it is difficult to use point features as control primitives. It is observed that most of the historical maps include areal features. Therefore, they tried to use areal features as control primitives to perform historical map to image registration. The technique is based on the shape context. The methodology is divided into 5 parts. The first part is to match polygons in a map space and geo-referenced coordinate system. The second part is to generate sample vertices in both polygons. The third part is to plot the log-polar histogram for every vertex. The fourth part is to match the log-polar histograms. The fifth part is to pair-up the vertices in map and geo-referenced spaces. The sixth part to perform coordinate transformation using matched vertices. Figure 2-3 shows the result of the experiment. Affine transformation and polynomial transformation were performed. It is observed that the use of a larger number of control polygons yield a better registration accuracy. In this dissertation, the method is similar to the approaches presented above. It will be explained in the next chapter.

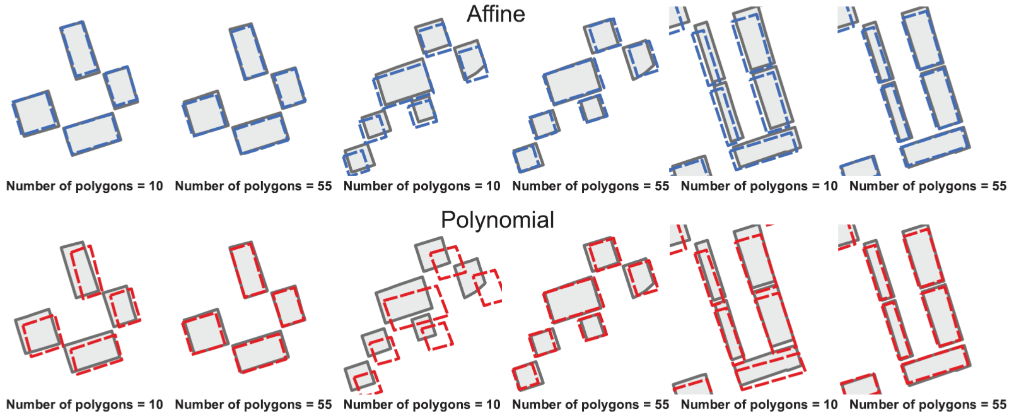


Figure 2-3: Result of polygon based image registration. Adapted from Yan, W., Easa, S., & Shaker, A. (2017). Polygon-based image registration: A new approach for geo-referencing historical maps. Remote Sensing Letters, 8(7), P.710.

Dice Similarity Coefficient (DSC) is a statistic used to estimate the similarity of two samples. It was known by several other names, including Sørensen–Dice index, Sørensen index and Dice's coefficient. Dice (1945) was the first person to introduce the algorithm. He introduced the coefficient of association of Forbes at the beginning of the article. It indicates the amount of association between two given species, compared to the amount of association between them. He stated that this algorithm was too complicated for general use. Then, he proposed a new algorithm named as association index. This was a simple direct measure of the amount of association of one species with another the association index. If α is the number of random samples of a given series in which species A occurs and h is the number of samples in which another species B occurs together with A, then the association index can be calculated by: -

$$Association\ index\ \frac{B}{A} = \frac{h}{a}$$

Equation 2-1

The same formula applied if b is the number of samples in which species B occurs. He also proposed a coincidence index: -

$$Coincidence\ index = \frac{2h}{a+b}$$

Equation 2-2

This value is intermediate between the two reciprocal association indices. Then, Chi-square test will be used for measuring the statistical reliability of the deviation shown by the samples of a given series from the amount of association expected by chance. In recent years, DSC has commonly used the medical image registration. Alterovitz et al. (2006) developed a new deformable image registration method to register diagnostic probe-in magnetic resonance images to therapeutic probe out MR images. The method was evaluated using two metrics: DSC and point error. The DSC was computed using the prostate outline in the probe-in image and the prostate outline in the deformed probe-out image. The formula is defined as: -

$$D = \frac{2a}{2a+b+c}$$

Equation 2-3

where a is the number of picture elements (pixels) shared by both areas, b is the number of pixels unique to the first area, and c is the number of pixels unique to the second area.

Yan et al. (2017) also use DSC as a method of evaluation in the geo-registration. From their experiment, a is the overlapping area found in the paired-up polygon after image registration, b is the area identified in the GIS polygon which is not overlapped with the corresponding polygon from the historical map, and c is the area of the polygon of the historical map not being overlapped with the corresponding GIS polygon. The DSC is a scalar ranging from 0 to 1. The higher the value, the better the quality registration, vice versa.

Chapter 3

Shape Context Descriptor

This chapter presents the approach for using shape context descriptor on the boundaries in two different survey record. The descriptor is based on the concept of polygon-based image registration (Yan, Easa and Shaker, 2017) that has been discussed (refer to Section 2.3). This dissertation targeted on one polygon and another polygon, not all the polygons on a map. The polygons in map space and geo-referenced coordinate system were matched. Sample vertices will be generated in both polygons. The log-polar histogram for every vertex was plotted. The plotted log-polar histograms were matched. The vertices in the map and geo-referenced spaces were paired up. The 2D affined transformation was performed using matched vertices. DSC was computed to assess the similarity of the two polygons.

3.1 Method

3.1.1 Overall Workflow

Figure 3-1 illustrates the overall workflow for shape context descriptor of the two polygons. In this study, there will be two polygons named as "Shape A" and "Shape B". Their geometric information is stored in the shapefile(.shp). Each of the shapes contains several vertices in the edge. If the vertices found along the edges of a polygon is insufficient, sample vertices can be generated. An interval between each of the vertices can be defined. The distance and angle for each of the vertex with respect to all the other vertices can be computed, using its x and y coordinates. The distance will be normalized since the scale of each of the shape may be different. The log-polar histogram will be plotted based on the normalized distance and angle. After that, the cost matrix for each of the histograms with respect to other histograms will be computed. Vertices between two shapes will be paired up based on their cost matrix. The 2D affine transformation will be performed to correlate the two shapes, using the paired-up vertices. Finally, DSC between the two shapes will be computed.

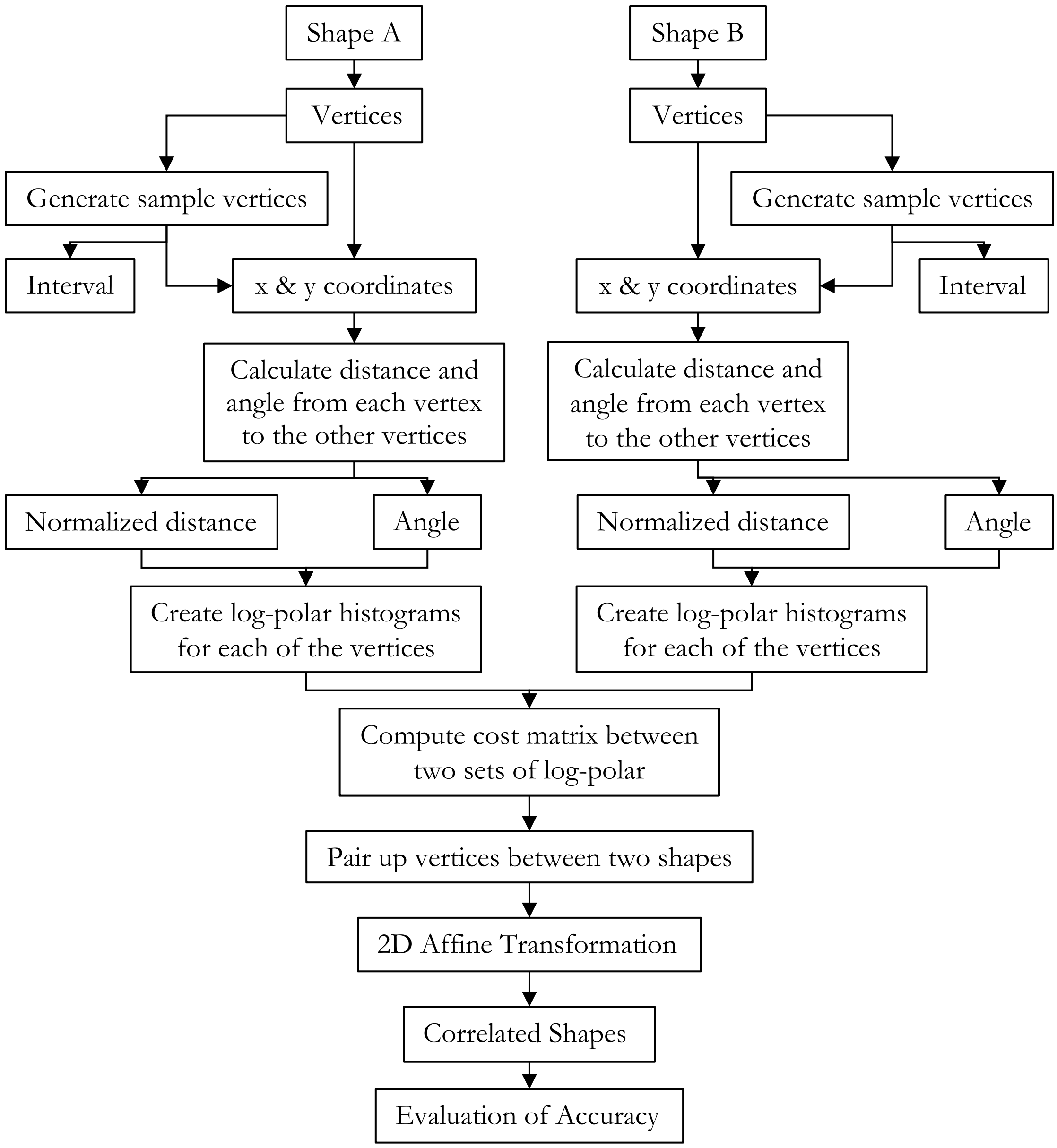


Figure 3-1: Overall workflow of Shape Context Descriptor

3.1.2 Polygons' Vertices Extraction

Assume a polygon P has k number of vertices v in a two-dimensional real space R^2. Any two adjacent vertices form an edge vector e of the polygon P. The polygon P is composed of:

$$\{v_1, \dots, v_i, v_j \dots, v_k\} \in R^2$$

Equation 3-1

where k > j > i > 1. v_1 is the first vertex v. v_i, v_j is the number i and j of the vertex v respectively. v_k is the last vertex of the polygon. $\{v_i, v_j\}$ is named as edge vector e, which will be further explained in the next paragraph. Figure 3-2 shows two triangles in a different size. The bigger triangle is named as Shape A while the smaller one is named as Shape B. Both A and B have 3 vertices as it is the least number of vertices to form a triangle. By Equation 3-1, $A = \{v_{A1}, v_{A2}, v_{A3}\} \in R^2$, and $B = \{v_{B1}, v_{B2}, v_{B3}\} \in R^2$.

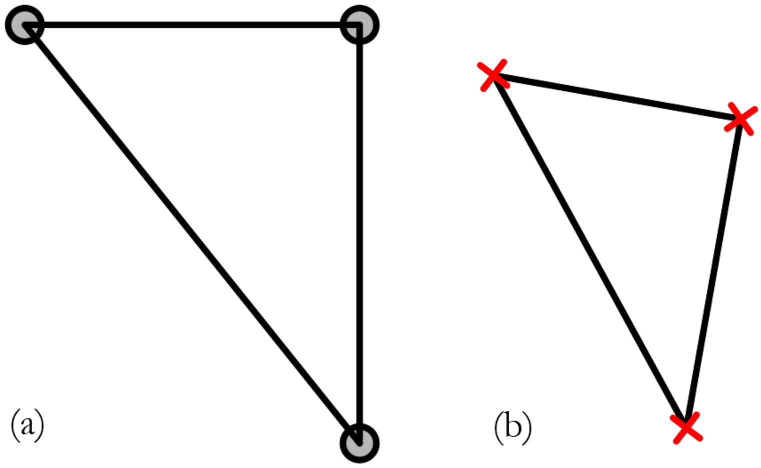


Figure 3-2: Two triangles in different size. (a) shows a triangle in bigger size, named as Shape A. (b) shows a triangle in smaller size, named as Shape B.

Figure 3-3 shows the script used to read the shapefile in Python 3.7. The geometric information of the two shapes are stored in the shapefile (.shp). The Python Shapefile Library (PyShp) was imported. A new "Reader" object was created to read the shapefile. The shapefile's geometry can be viewed by calling the shapes( ) method.

```
import shapefile

def read_shapefile(shape_path):
    sf = shapefile.Reader(shape_path)
    with shapefile.Reader(shape_path) as shp:
        print(shp)
    print(sf.shapeType) #1 is Point, 5 is polygon
    shapes = sf.shapes()
    return shapes

def get_xy(shapes):
    x=[]
    y=[]
    xy=[]
    for i in range (0, len(shapes[0].points)):
        shape = shapes[0].points[i]
        xy_object = ['%.3f' % coord for coord in shape]
        x.append(float(xy_object[0]))
        y.append(float(xy_object[1]))
        xy.append(float(xy_object[0]))
        xy.append(float(xy_object[1]))
    return xy
```

Figure 3-3: Script of reading shapefile in Python 3.7

To perform the shape context descriptor, the number of vertices should be considered. If the vertices along the edges of a polygon are insufficient, the performance of shape context descriptor will be affected. In other words, the smaller the interval between the edge vector e (i.e. l(e) is small, where l refers to the length of the edge), the more sample vertices will be generated. However, more computational resources are required at the same time. An interval d can be specified. Yan, Easa and Shaker (2017) suggested that a map with scale 1:5000 can set the interval as 5. Then a [l(e)/d] number of sample vertices can be generated along with each e. There is a total number of K vertices after adding sample vertices. Figure 3-4 shows the triangles after adding sample vertices. The interval d between each adjacent vertex v is fixed.

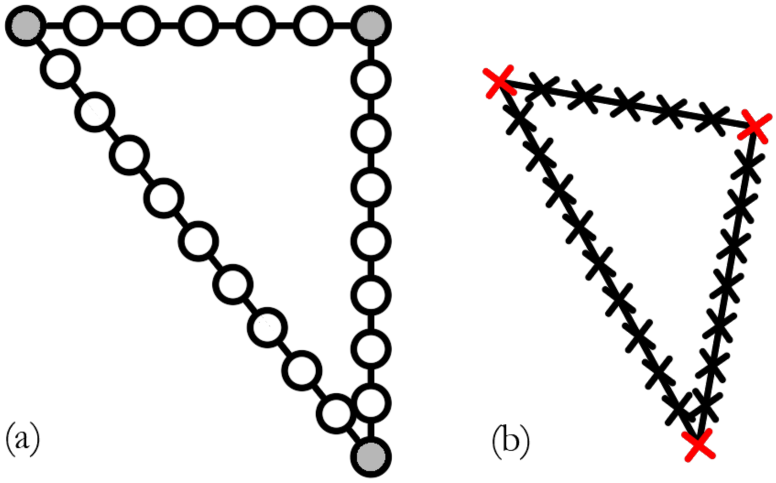


Figure 3-4: The triangles after adding sample vertices

Figure 3-5 shows the script for adding sample vertices. The interval d was set as 4. When the interval between two adjacent vertices is larger than 4, a sample vertex will be added in the mid-point of the edge.

```
import shapefile

def add_dummy (xy,dist,dist_index):
    i = dist_index*2
    mid_x=(xy[i]+xy[i+2])/2
    mid_y=(xy[i+1]+xy[i+3])/2
    xy.insert((i+2),mid_x)
    xy.insert((i+3),mid_y)
    return xy

#adding dummy points
n=0
while n==0:
    dist=[]
    for i in range (0,len(xy)-3,2):
        dist.append(abs( math.sqrt((xy[i]-xy[i+2])**2 + (xy[i+1]-xy[i+3])**2) )
        max_pos=dist.index(max(dist))
    if max(dist) > 4: #input the distance here
        xy=add_dummy(xy,dist,max_pos) #calling a function
    else:
        break
```

Figure 3-5: Script of adding sample vertices in Python 3.7

3.1.3 Log-polar Histograms

Each of the K vertices in Polygon P has an x and y coordinates. The logarithmic distance (D) and whole circle bearing (θ) for each vertex with respect to all the other K – 1 vertices, will be calculated using the x and y coordinates.

$$d_{v_i - v_j} = ln\sqrt{\left(x(v_i) - x(v_j)\right)^2 + \left(y(v_i) - y(v_j)\right)^2} \quad where\ \forall i, j \in [1, K] \mid i \neq j$$

Equation 3-2

$$\theta_{v_i - v_j} = arctan\left[\frac{y(v_j) - y(v_i)}{x(v_j) - x(v_i)}\right] \quad where\ x(v_i) \neq x(v_j) \mid \forall i, j \in [1, K] \mid i \neq j$$

Equation 3-3

Taking Shape A as an example (Figure 3-6 a). The formula to calculate the distance (Equation 3-4) and whole circle bearing (Equation 3-5) of vertex A1 with respect to the other 23 vertices are listed as follow:

$$d_{v_{A1} - v_{A2}} = ln\sqrt{\left(\mathrm{x}(\mathrm{v}_{\mathrm{A1}}) - \mathrm{x}(\mathrm{v}_{\mathrm{A2}})\right)^2 + \left(\mathrm{y}(\mathrm{v}_{\mathrm{A1}}) - \mathrm{y}(\mathrm{v}_{\mathrm{A2}})\right)^2}$$
$$\mathrm{d}_{\mathrm{v}_{\mathrm{A1}} - \mathrm{v}_{\mathrm{A3}}} = ln\sqrt{\left(\mathrm{x}(\mathrm{v}_{\mathrm{A1}}) - \mathrm{x}(\mathrm{v}_{\mathrm{A3}})\right)^2 + \left(\mathrm{y}(\mathrm{v}_{\mathrm{A1}}) - \mathrm{y}(\mathrm{v}_{\mathrm{A3}})\right)^2}$$
$$\vdots$$
$$\mathrm{d}_{\mathrm{v}_{\mathrm{A1}} - \mathrm{v}_{\mathrm{A24}}} = ln\sqrt{\left(\mathrm{x}(\mathrm{v}_{\mathrm{A1}}) - \mathrm{x}(\mathrm{v}_{\mathrm{A2}})\right)^2 + \left(\mathrm{y}(\mathrm{v}_{\mathrm{A1}}) - \mathrm{y}(\mathrm{v}_{\mathrm{A2}})\right)^2}$$

Equation 3-4

$$\theta_{\mathrm{v}_{\mathrm{A1}} - \mathrm{v}_{\mathrm{A2}}} = \mathrm{arctan}\left[\frac{\mathrm{y}(\mathrm{v}_{\mathrm{A2}}) - \mathrm{y}(\mathrm{v}_{\mathrm{A1}})}{\mathrm{x}(\mathrm{v}_{\mathrm{A2}}) - \mathrm{x}(\mathrm{v}_{\mathrm{A1}})}\right]$$
$$\theta_{\mathrm{v}_{\mathrm{A1}} - \mathrm{v}_{\mathrm{A3}}} = \mathrm{arctan}\left[\frac{\mathrm{y}(\mathrm{v}_{\mathrm{A3}}) - \mathrm{y}(\mathrm{v}_{\mathrm{A1}})}{\mathrm{x}(\mathrm{v}_{\mathrm{A3}}) - \mathrm{x}(\mathrm{v}_{\mathrm{A1}})}\right]$$
$$\vdots$$
$$\theta_{\mathrm{v}_{\mathrm{A1}} - \mathrm{v}_{\mathrm{A24}}} = \mathrm{arctan}\left[\frac{\mathrm{y}(\mathrm{v}_{\mathrm{A24}}) - \mathrm{y}(\mathrm{v}_{\mathrm{A1}})}{\mathrm{x}(\mathrm{v}_{\mathrm{A24}}) - \mathrm{x}(\mathrm{v}_{\mathrm{A1}})}\right]$$

Equation 3-5

The process is then repeated for vertex A2, shown in the Figure 3-6 (b).

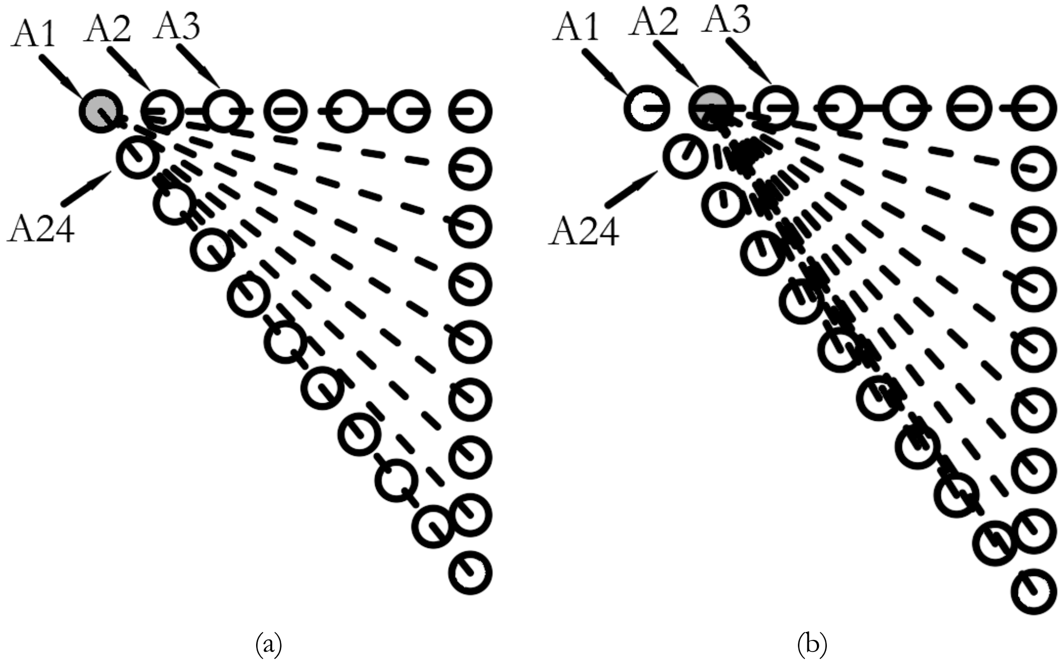


Figure 3-6. (a) Computation of vertex A1 with respect to the other 23 vertices. (b) Computation of vertex A2 with respect to the other 23 vertices

If the two polygons are not in the same coordinate system, the normalization of the distance should be performed. For each shape, the distance should be normalized with respect to its median value. The computed distance and angles will be stored in a K-1 by 2 matrices.

$$v_i(D,\theta) = \begin{bmatrix} D_{v_i-v_1} & \theta_{v_i-v_1} \\ \vdots & \vdots \\ D_{v_i-v_j} & \theta_{v_i-v_j} \\ \vdots & \vdots \\ D_{v_i-v_{K-1}} & \theta_{v_i-v_{K-1}} \end{bmatrix}$$

Equation 3-6

The normalized distance and whole circle bearing will be used to form a log-polar histogram H with a bin size b_D and b_θ for p and θ respectively. b_D and b_θ are based on the size of histogram, B_D by B_θ. The data structure of a log-polar histogram H is defined as follow:

$$H = \begin{bmatrix} N(B_D b_D, b_\theta) & \cdots & N(D_D b_D, B_\theta b_\theta) \\ \vdots & \ddots & \vdots \\ N(b_D, b_\theta) & \cdots & N(b_D, B_\theta b_\theta) \end{bmatrix}$$

Equation 3-7

Where N is the number of occurrences in a specific bin found in the log-polar histogram. If the size of histogram $(B_D \times B_\theta)$ is set as 5 by 12, the bin size b_D and b_θ will be 0.4 and 30° respectively. If Shape A has $K^{(A)}$ vertices, there will be $K^{(A)}-1$ log-polar histogram generated. It is noted number of vertices $K^{(B)}$ is not necessarily equal to $K^{(A)}$.

The data structure can be further explained in Figure 3-7. It is the log-polar histogram of

Shape A. In the left bottom corner, it is the data bin of N(1*0.4, 1*30°) = N(0.4, 30°). Results within 0 – 0.4 distance, and 0° - 30°, will be put into this data bin of. The darkest bin, which means it has the highest occurrences (i.e. 4 occurrences), is N(4*0.4, 2*30°) = N(1.6, 60°). Results within 1.2 – 1.6 distance and 30° - 60° will be put into this data bin.

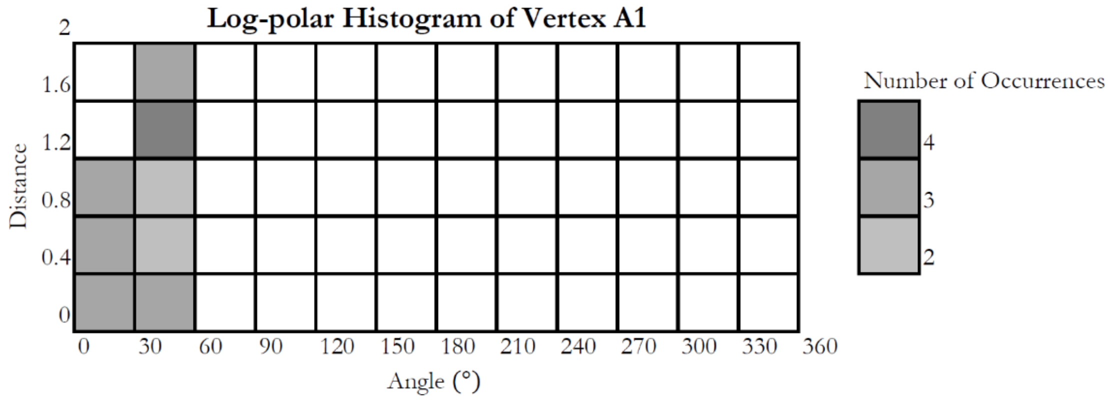


Figure 3-7: Log-polar histogram of Vertex A1 of Shape A

Figure 3-8 shows the script of plotting the log-polar histograms. The size of the log-polar histogram is 12 * 5, which is 60 grids. A list with 60 of “0” representing the log-polar histogram is created. Then, “for” and “if” loop are used to plot the diagram. Distance and whole circle bearing are imported pair by pair. Distance will be checked first. If the distance is within 0 and 0.4, then the next step is to check the angle. If not, the checking on distance will be continued by adding 0.4. Same theory applies on the angle checking. After plotting the lop-polar histogram of that vertex, the result will be exported to a .data file.

```python
def log_polar(dist_angle):
    log_po=[]
    for i in range(60):
        log_po.append(0)
    # 12*5 log-polar historgram, totally 60 grid
    for count in range(0,len(dist_angle),2):
        d=0 # 0-2, 0.4
        a=0 # 0-360, 30
        for yy in range (0,5):
            if d<dist_angle[count]<d+0.4:
                for xx in range (0,12):
                    if a<dist_angle[count+1]<a+30:
                        index = grid(yy,xx)
                        log_po[index]=log_po[index]+1
                        break
                    else:
                        a=a+30
                break
            else:
                d=d+0.4
    return log_po

def grid(y,x):
    y_grid = 0
    for i in range(0,60,12):
        if y == y_grid:
            pos=i+x
            break
        else:
            y_grid=y_grid+1
    return pos

# input the list of distance and angle into the log polar
for i in range (1,int(len(xy)/2)):
    path_r="output/A/A{}.data"
    with open(path_r.format(i), 'rb') as filehandle:
        # read the data as binary data stream
        A_dist_angle = pickle.load(filehandle)
        log_po_hist=log_polar(A_dist_angle)
        path_w="output/A/A{}_log-polar_histogram.data"
        with open(path_w.format(i), 'wb') as filehandle:
            # store the data as binary data stream
            pickle.dump(log_po_hist, filehandle)
```

Figure 3-8: Script of plotting log-polar histogram in Python 3.7

3.1.4 Vertices Matching

To match the vertices of Shape A and those of Shape B, cost matrix C between two log-polar histograms will be calculated.

$$\mathrm{C} = \frac{1}{2}\sum_{k_H=1}^{K_H}\frac{[\mathrm{A}(k_H) - \mathrm{B}(k_H)]^2}{\mathrm{A}(k_H) + \mathrm{B}(k_H)}$$

Equation 3-8

Where K_H is the total number of data bins in the histogram, k_H is the specific position of data bin within the histogram, A and B are the log-polar histogram of Shape A and Shape B respectively. For each vertex v in Shape A, $K^{(B)}$ of cost matrix will be calculated. Each cost matrix represents the correlation between the subjective vertex and that compared vertex. Here is an example of the computation of cost matrix C (Figure 3-9). The four lower left grids are extracted from the histogram of the first vertex of Shape A and the histogram of the first vertex of Shape B respectively. The cost matrix is calculated in Equation 3-9.

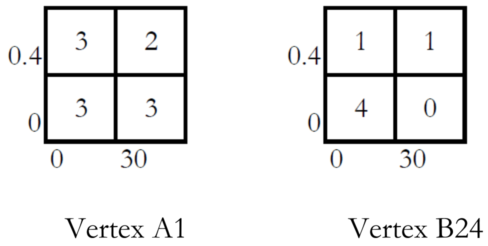


Figure 3-9: Example of cost matrix's calculation

$$C = \frac{1}{2} \times \left[\frac{(3-4)^2}{3+4} + \frac{(3-0)^2}{3+0} + \frac{(3-1)^2}{3+1} + \frac{(2-1)^2}{2+1} + \cdots \right]$$
$$= 2.238 + \cdots$$

Equation 3-9

Assume there are 24 vertices in Shape B, 24 cost matrices will be computed. The cost matrix with the minimum value will be paired up. The detail can be further illustrated in Equation 3-10. The minimum C is 0.45, which means that vertex A1 will be paired up with vertex B2.

$$\begin{gathered} A1 - B1 \rightarrow C = 0.21 \\ A1 - B2 \rightarrow C = 0.45 \\ \vdots \\ A1 - B24 \rightarrow C = 2.89 \end{gathered}$$

Equation 3-10

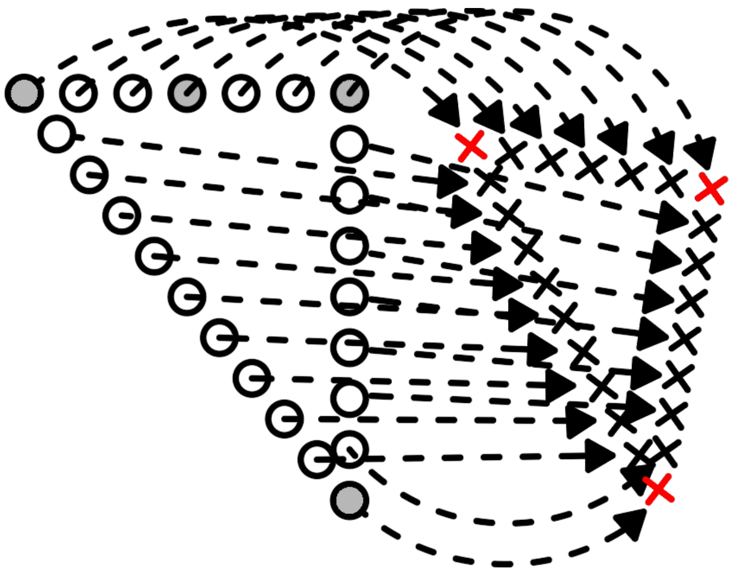


Figure 3-10: Pair up the vertices between two shapes

Figure 3-11 shows the code of performing vertices matching between the two shapes. After importing the data of log-polar histograms of the two shapes, the cost matrix of each pair of vertices will be computed. Each pair of matched vertices will be recorded by a list.

```
def cost_mat(A,B):
    fraction = 0
    for i in range(len(A)):
        if A[i]==0 & B[i]==0:
            continue
        else:
            fraction = fraction + ( ((A[i]-B[i])**2) / (A[i]+B[i]) )
    ans = fraction/2
    return ans

path_r="output/AB/A_coord.data" # the first shape
with open(path_r, 'rb') as filehandle:
    # read the data as binary data stream
    A_xy = pickle.load(filehandle)
A_xy.pop(len(A_xy)-1)
A_xy.pop(len(A_xy)-1)

path_r="output/AB/B_coord.data" # the second shape
with open(path_r, 'rb') as filehandle:
    # read the data as binary data stream
    B = pickle.load(filehandle)
B.pop(len(B)-1)
B.pop(len(B)-1)

B_xy=[]
C_list=[]
for i in range(1,int(len(A_xy)/2)+1):
    C=[]
    path_r="output/AB/A{}_log-polar_histogram.data"
    with open(path_r.format(i), 'rb') as filehandle:
        # read the data as binary data stream
        A_l_p = pickle.load(filehandle)
    for j in range(1,int(len(B)/2)+1):
        path_r="output/AB/B{}_log-polar_histogram.data"
        with open(path_r.format(j), 'rb') as filehandle:
            # read the data as binary data stream
            B_l_p = pickle.load(filehandle)
            C.append(cost_mat(A_l_p,B_l_p))
    mini_pos = C.index(min(C))+1
    C_list.append(min(C))
    B_xy.append(B[C.index(min(C))*2]) # store the matched vertices
    B_xy.append(B[C.index(min(C))*2+1])
```

Figure 3-11: Script of performing vertices matching in Python 3.7

3.1.5 Coordinate Transformation

After the matching of vertices between the two shapes, coordinate transformation can be performed. 2D Affine transformation is chosen in this experiment.

$$X = Ax + By + C$$

Equation 3-11

$$Y = Dx + Ey + F$$

Equation 3-12

Where X and Y are the coordinates of points in the geo-referenced coordinate system. x and y are the coordinates of points in the map coordinate system. It depends on which shape is the geo-referenced coordinate system or map coordinate system: Shape A and Shape B, vice versa. A, B, C, D, E F are the transformation parameters needed to be determined. At least 3 pairs of control points are required to solve the 6 parameters. The least squares adjustment will be used in the computation (Equation 3-13).

$$l = \beta x$$

Equation 3-13

where elements of vector l are the results of the geo-referenced coordinate system, which are related to the elements of unknown vector x. x contains the results of map coordinate system. β is called the coefficient matrix, which is composed of A to F. Assuming n pair of vertices are used as the control points, l will be an array with size of 2n by 1. A will be a matrix with size of 2n by 6.

It is assumed that Shape A is in the geo-referenced coordinate system while Shape B is in the map coordinate system. Vertex A1 is paired up with B1, A2 is paired up with B2, and A3 is paired up with B3. 3 pair of vertices are used as the control points.

$$\text{Observation Matrix} = l = \begin{bmatrix} X_{A1} \\ Y_{A1} \\ X_{A2} \\ Y_{A2} \\ X_{A3} \\ Y_{A3} \end{bmatrix}$$

Equation 3-14

$$\text{Design Matrix} = \beta = \begin{bmatrix} x_{B1} & y_{B1} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{B1} & y_{B1} & 1 \\ x_{B2} & y_{B2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{B2} & y_{B2} & 1 \\ x_{B3} & y_{B3} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{B3} & y_{B3} & 1 \end{bmatrix}$$

Equation 3-15

$$\text{Parameter Matrix} = x = \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix}$$

Equation 3-16

Equation 3-13 becomes:

$$l = \beta x$$

$$\begin{bmatrix} X_{A1} \\ Y_{A1} \\ X_{A2} \\ Y_{A2} \\ X_{A3} \\ Y_{A3} \end{bmatrix} = \begin{bmatrix} x_{B1} & y_{B1} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{B1} & y_{B1} & 1 \\ x_{B2} & y_{B2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{B2} & y_{B2} & 1 \\ x_{B3} & y_{B3} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{B3} & y_{B3} & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix}$$

Equation 3-17

The coefficient matrix can be calculated by:

$$x = (A^{T}A)^{-1}(A^{T}l)$$

Equation 3-18

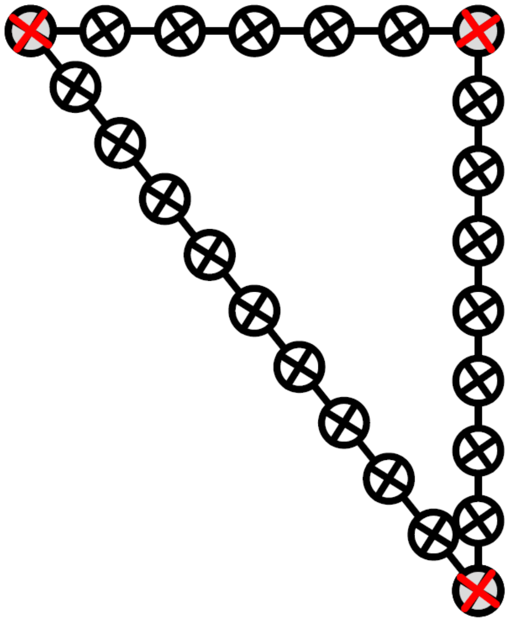


Figure 3-12: Shape A and Shape B after performing coordinate transformation

Figure 3-13 shows the script of performing the 2D affine transformation. In the previous part (Figure 3-11), the result of vertices matching was stored in a list. In this part, the x and y coordinates of each paired-up vertices will be collected and put into the observation matrix and parameter matrix. NumPy is used, which is a package for scientific computing which supports for a powerful N-dimensional array object.

```
#copy a list of DD coordinates for affine transformation
A_xy_DD=[]
for i in range(0,len(A_xy)):
    A_xy_DD.append(A_xy[i])

# Affine Transformation, l=Ax
l=[]
A=[]
for i in range (0,10):
    min_index = C_list.index(min(C_list))
    C_list.pop(min_index)
    l.append([B_xy.pop(2*min_index)]) #l-->SRP
    l.append([B_xy.pop(2*min_index)])
    x_temp = A_xy.pop(2*min_index)
    y_temp = A_xy.pop(2*min_index)
    A.append([x_temp, y_temp, 1, 0, 0, 0]) #A-->DD
    A.append([0,0,0,x_temp,y_temp,1])

l=np.matrix(l)
A=np.matrix(A)

x= ( ( (A.T).dot(A) ).I ).dot( (A.T).dot(l) )
```

Figure 3-13: Script of performing the 2D affine transformation

3.1.6 Dice Similarity Coefficient

For the similarity assessment of the two shapes, DSC will be measured.

$$DSC = \frac{2A}{2A + B + C}$$

Equation 3-19

Where A is the overlapping area found in the paired-up polygon after the coordinate transformation, B is the area identified in the shape from the geo-referenced coordinate system which is not overlapped with the corresponding shape from the map coordinate system, C is the area identified in the shape from the map coordinate system which is not overlapped with correspond shape from the geo-referenced coordinate system. The theory of DSC is similar to the Venn Diagram (Figure 3-14). DSC can be computed in ArcMap. Under the "Geoprocessing", "Intersect" function can be performed between two polygons shapefile. The overlapping area can be computed. By "Field Calculator" in "Attribute Table", Equation 3-19 was input to compute DSC. DSC ranges from 0 to 1. A DSC value of 1 represents the two shapes are completely overlapped after registration, while a value of 0 represents they are completely off from each other. A higher DSC indicates the two shapes are similar to each other, vice versa.

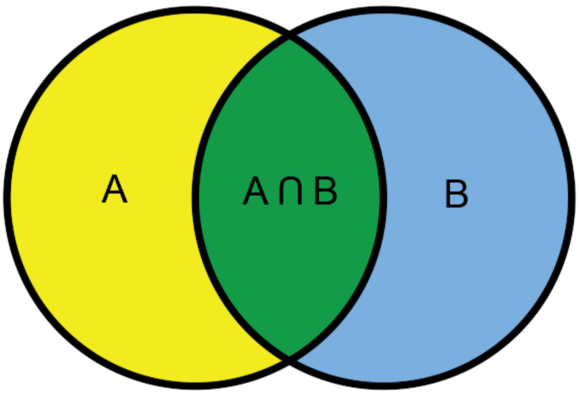


Figure 3-14: Shading Venn Diagram (A intersect B)

3.2 Experiment Testing

3.2.1 Study Area and Dataset

The study area covers 2 Demarcation District (DD): DD 1TC and DD 129. DD 1 TC was plotted at 1:3960. It depicted the lots in Tung Chung (Figure 3-15). Figure 3-16 shows the current location of DD 1 TC and the depicted lots in Tung Chung. DD 129 was plotted at 1:1980. The lots were depicted in Lau Fan Shan (Figure 3-17 & Figure 3-18). The current location of DD 120 is shown in Figure 3-19.

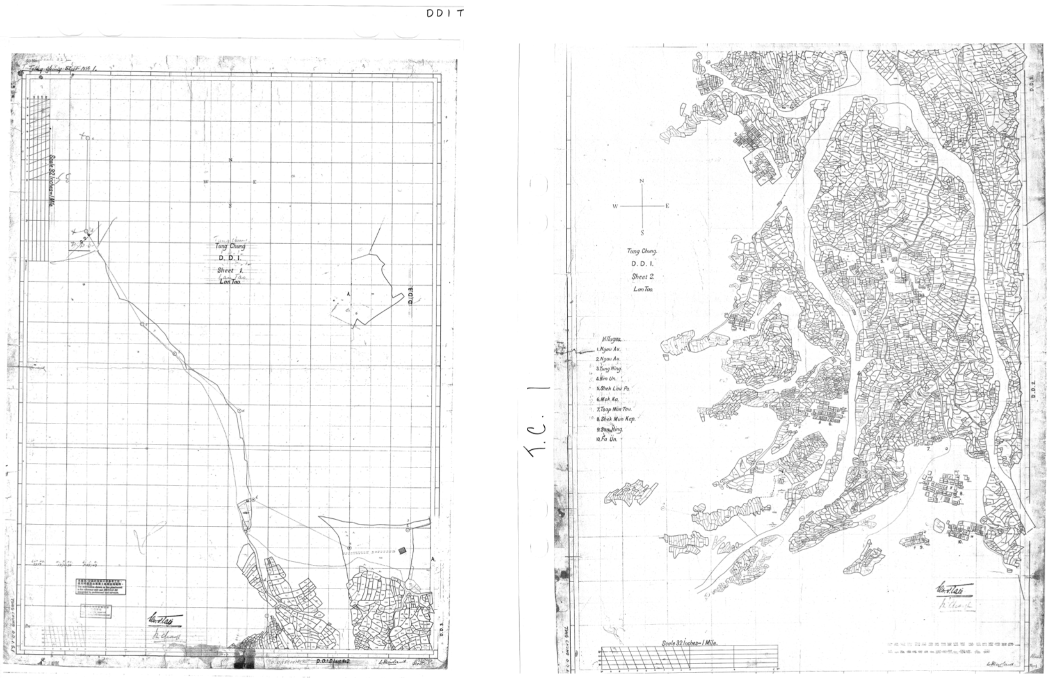


Figure 3-15: DD Sheet (Part 1 & Part 2 of DD 1 TC)

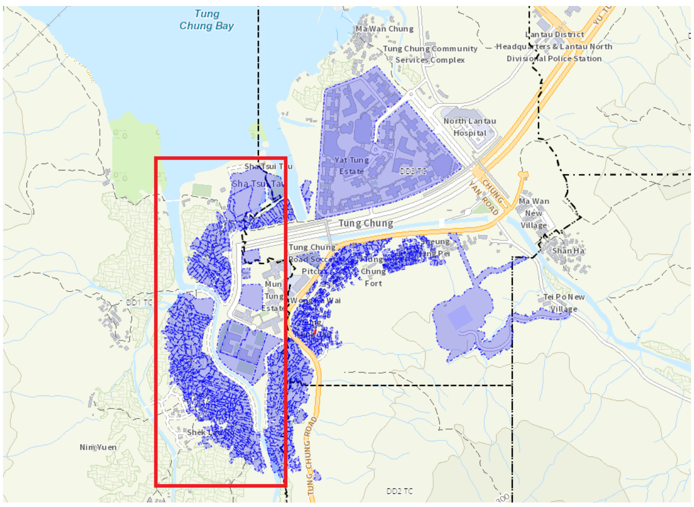


Figure 3-16: Location of DD 1 TC shown in HKMS 2.0 (the area inside the red bounding

box)

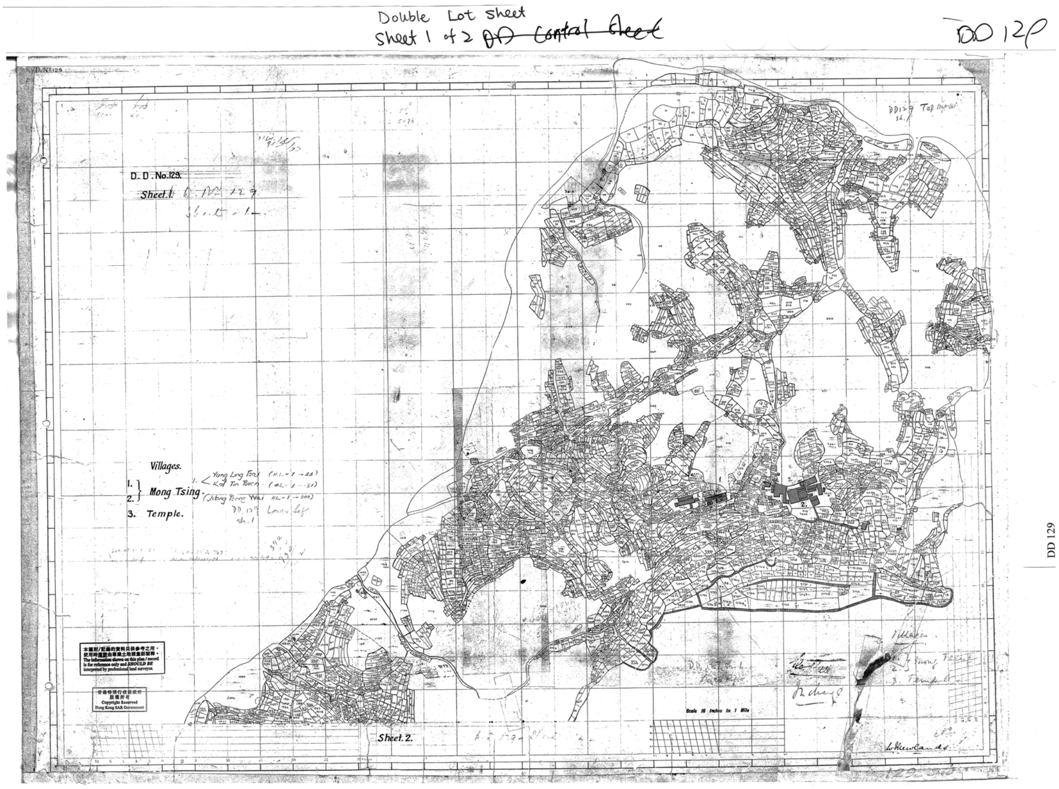


Figure 3-17: DD Sheet (Part 1 of DD 129)

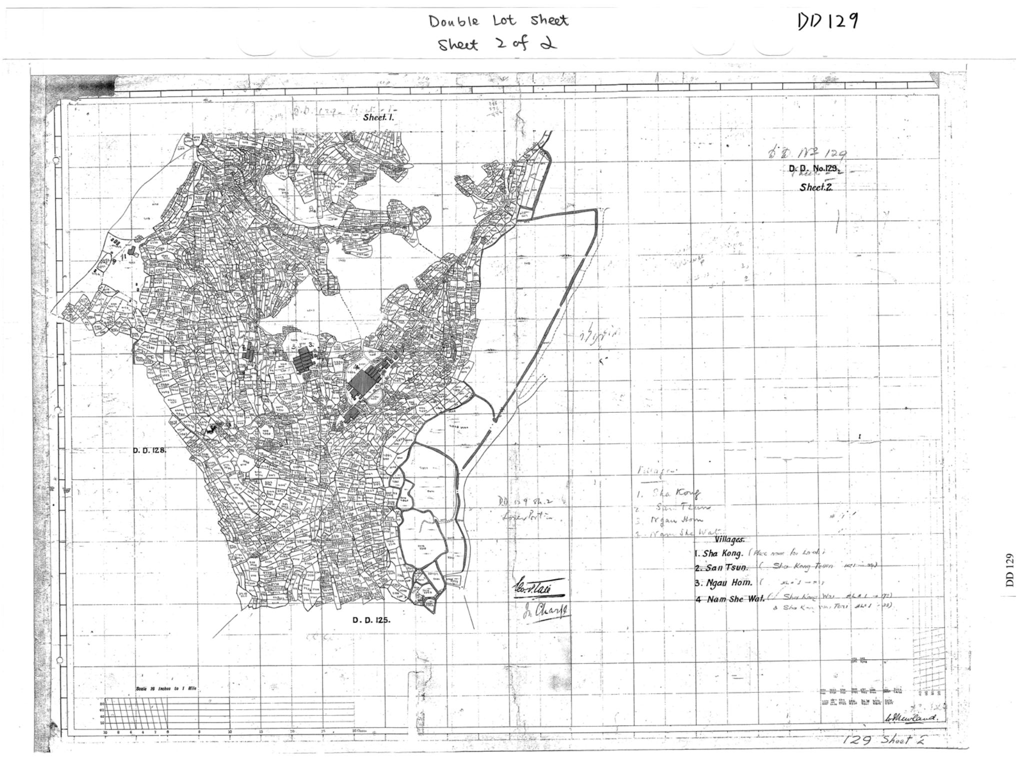


Figure 3-18: DD Sheet (Part 2 of DD 129)

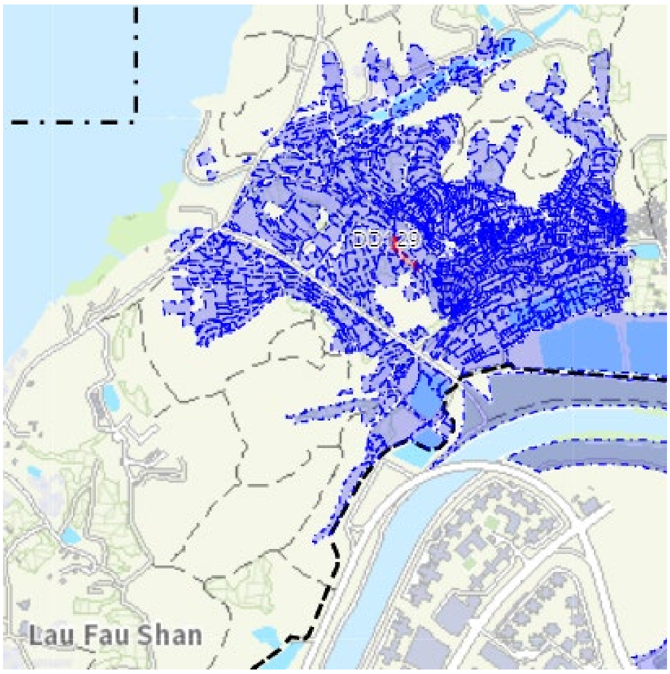


Figure 3-19: Location of DD 129 shown in HKMS 2.0

Totally 40 lots were selected as the sample data. 20 lots were found from DD 1 TC and 20 lots were found from DD 129. They were all undergone boundary re-establishment. Totally 40 survey record plan (SRP) were collected. Some samples data are shown below. Figure 3-20 is the Schedule Book of lot 1281 - 1311. Figure 3-21 is the SRP of lot 1392 in DD 129.

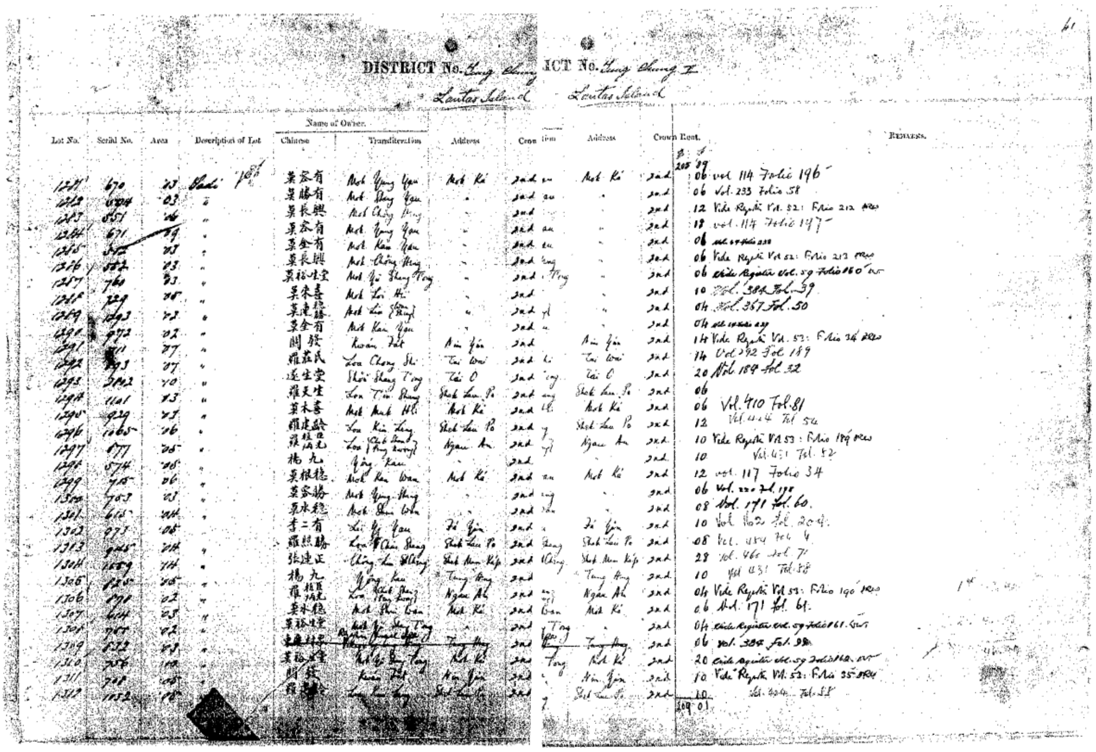


Figure 3-20: Schedule Book of lot 1281 – 1311

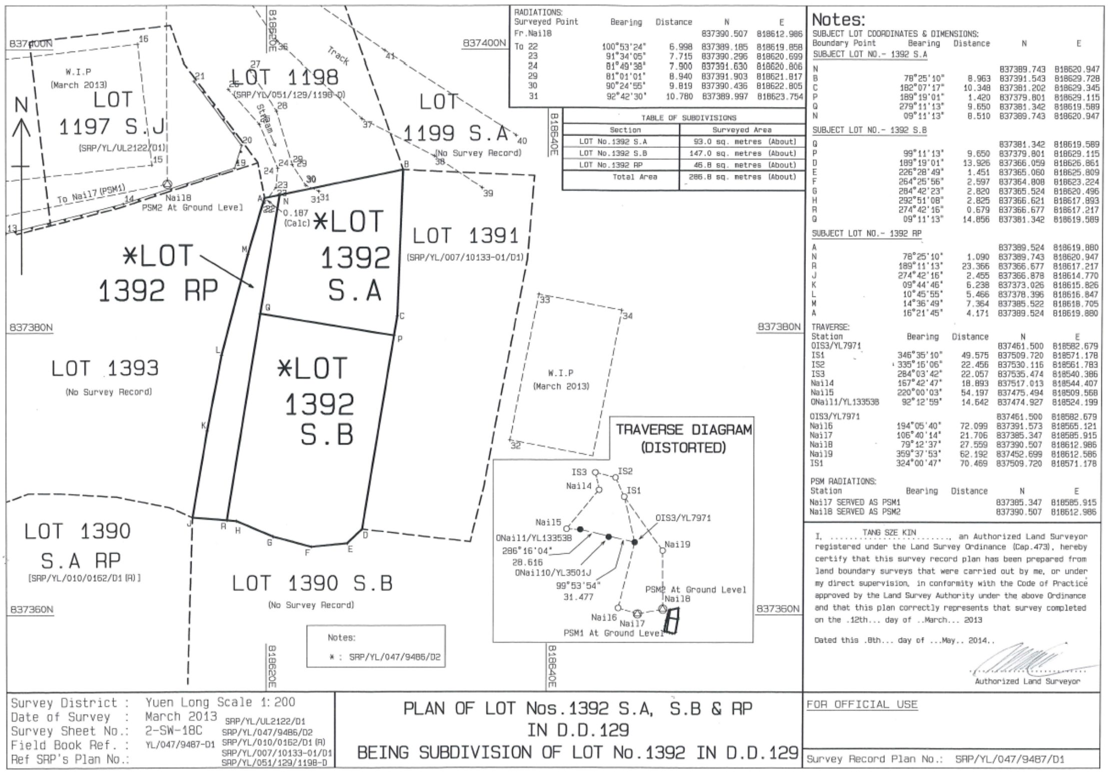


Figure 3-21: The SRP of Lot 1392 in DD 129

3.2.2 Design of Experiment and Evaluation

Following the method in Section 3.1, shape context was used for the correlation between Demarcation District (DD) sheet and survey record plan (SRP). DD sheet is identified in the map coordinate system while SRP is identified in the geo-referenced coordinate system. ArcMap 10.4.1 will be used to process the boundaries while Python 3.7 will be used for the computation. The procedures are stated below: -

1. The boundaries of selected lots in DD sheet were digitised as polygons;

2. The shapefiles of the boundaries were imported into Python 3.7;

3. For each boundary, sample vertices were added between two adjacent vertices in each edge;

4. Distance and whole circle bearing between one vertex with respect to other vertices were computed;

5. The computed distances were normalised by the median of the distance;

6. The nominalised distance and angle were store as a matrix;

7. The log-polar histogram of each vertex was plotted using its respecting matrix;
8. Procedure 1 - 7 were repeated on the SRP;
9. After generating the log-polar histograms, the cost matrix of each vertex of DD Sheet's boundary with respect to the vertices of SRP's boundary was calculated;
10. The vertices in DD Sheet's boundaries were paired up with those in SRP's boundaries;
11. 2D affine transformation was performed using the paired-up vertices;
12. Dice Similarity Coefficient (DSC) was computed to evaluate the similarity of two shapes

Some requirements were set in the procedures above. One of the requirements is the type of the selected lots should be the same. The surveying and demarcation technique may vary with land types. For example, the house lot was granted with the wall, while the agriculture land was granted with the field bund. The surveyed features were not the same. Therefore, same types of lots should be selected so a fair experiment can be conducted.

For the generation of sample vertices, different intervals were set for the two DD sheets. The interval was set as 1 in DD 1 TC. It means that the sample vertices will be added at least 1 m between two adjacent vertices. 1 m is chosen because it is the graphical accuracy of a DD sheet plotted at 1: 1980. In a court case related to land boundary dispute (Lam Pak Kau v. Yu Yuet Fat [1996] No. A2020, 3), the land surveyor hired by the defendant explained that it was impossible to place a boundary with more accuracy than within about 1 metre using the DD sheet at 1:1980, which was agreed by the court. Base on the standard stated above, the interval will be set as 2 for DD 129 as it was plotted at 1:3960. The sample vertices will be added at least 2 m between two adjacent vertices.

After calculating DSC, the values will be compared with the comment in the survey reports. For example, DSC of a lot (named as Lot A) is calculated. This value indicates the similarity of Lot A's boundary in DD sheet between Lot A's boundary in Survey Record Plan (SRP). The land surveyor will record the evidence used and procedure of the lot boundary re-establishment in the survey report. He or she will give comment on the consistency between the re-established boundary and DD sheet's boundary, in terms of shape and area. If the comment about the shape is consistent and the DSC is close to 1, this can prove that the DSC is valid to use.

Chapter 4

Result and Analysis

This chapter presents the result of performing shape context on the 40 lots depicted in DD 1 TC and DD 129. The methodology of shape context descriptor was explained in 10. There are two main issues to figure out. The first issue is to analyze the feasibility of DSC in assisting the lot boundary determination. To solve this issue, the digitised area of each lot in Demarcation District (DD) sheet will be correlated with the respecting surveyed area in Survey Record Plan (SRP). Dice Similarity Coefficient (DSC) will be calculated. The DSC will be then compared to the comment in the survey report. The second issue is to check the performance of shape context in two different scale maps. DD 1 TC was plotted at 1:1980 while DD 129 was plotted at 1:3960. The purpose is to observe if DSC will be similar or not, in DD 1 TC and DD 129.

4.1 Types of Selected Lots

Totally 40 lots were selected as the sample for the experiment, including 20 lots in DD 1 TC and 20 lots in DD 129. The land type of each lot can be found in its respecting Schedule book. In the red bounding box of Figure 4-1, "Description of Lot" stated the land use at the time of DD survey. "Padi" means paddy land. It was found that all the selected lots are the agriculture lot (Table 4-1.).

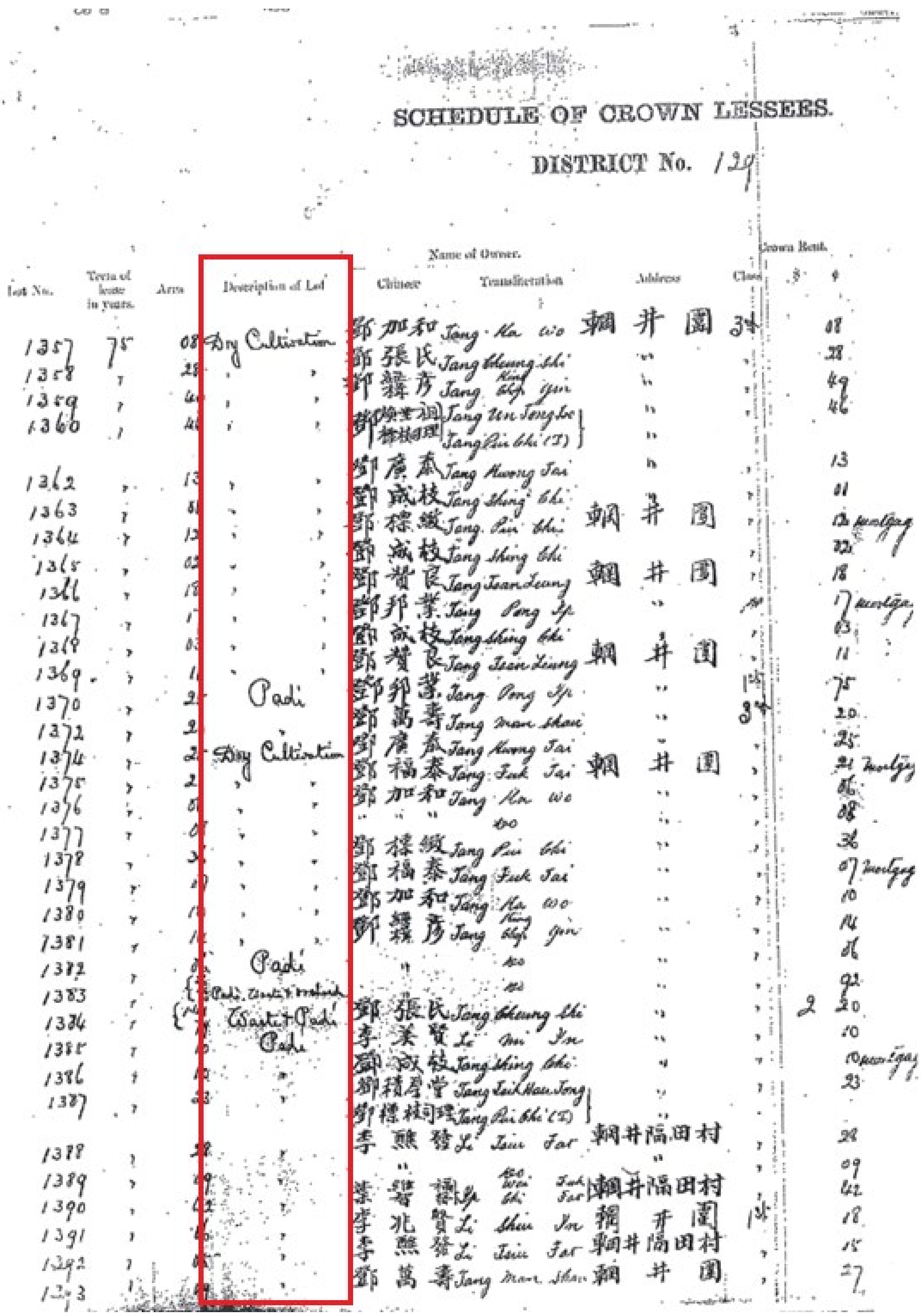


Figure 4-1: Schedule Book of Lot 1357 – 1393 in DD 1 TC

Table 4-1: Land Types of the 40 Selected Lots

DD 1 TC		DD 129	
Lot	**Type**	**Lot**	**Type**
1311	Padi	1157	Padi
1765	Padi	1158	Padi
2237	Padi	1180	Padi
2248	Padi	1181	Padi
2254	Padi	1196	Padi
2255	Padi	1198	Padi
2257	Padi	1201	Padi
2593	Padi	1218	Padi
2595	Padi	1353	Dry Cultivation
2597	Padi	1354	Dry Cultivation
2603	Padi	1357	Dry Cultivation
2627	Padi	1359	Dry Cultivation
2629	Padi	1376	Dry Cultivation
2631	Padi	1377	Dry Cultivation
2632	Padi	1380	Dry Cultivation
2633	Dry Cultivation	1388	Padi
2636	Dry Cultivation	1389	Padi
2643	Padi	1392	Padi
2648	Padi	1395	Padi
2651	Dry Cultivation	1419	Padi

4.2 Geo-referencing

Geo-referencing was conducted to transform the DD sheet into an actual scale. Also, the paper must be stretched since it has been used for a century. Geo-referencing can minimise error caused by paper stretching and scanning. To perform the geo-referencing, the actual grid length g was computed. The control points were added to the corner of the grids according to the grid length. For example, a corner of grid was set as (0,0). The corner next to it will be (g, 0) or (0, g). At least three non-collinear points should be added as the control points. Moreover, they should be added near the selected lots to minimise the effect of paper stretching, thus attain higher accuracy.

4.2.1 DD 1 TC

DD 1 TC was plotted at 32 inches to 1 mile, which is 1:1980. A gridline in the DD sheet has a length of 1 inch. To turn the length of a gridline into actual scale, in meter unit, the following approaches were conducted: -

1. 32 inches to 1 mile = 32 inch to 1609.344 m
2. 32 inches to 1609.344 m → 1 inch to 50.292 m

A gridline in a 1:1980 DD sheet equals to 50.292 m. The control points were added to the corner of grid based on the grid length of 50.292 m. Figure 4-2 shows the geo-referencing of DD 1 TC. The points in red are the selected 20 lots. The red cross are the control points. The total Root Mean Square (RMS) error was 0.297909 (Figure *4-3*)

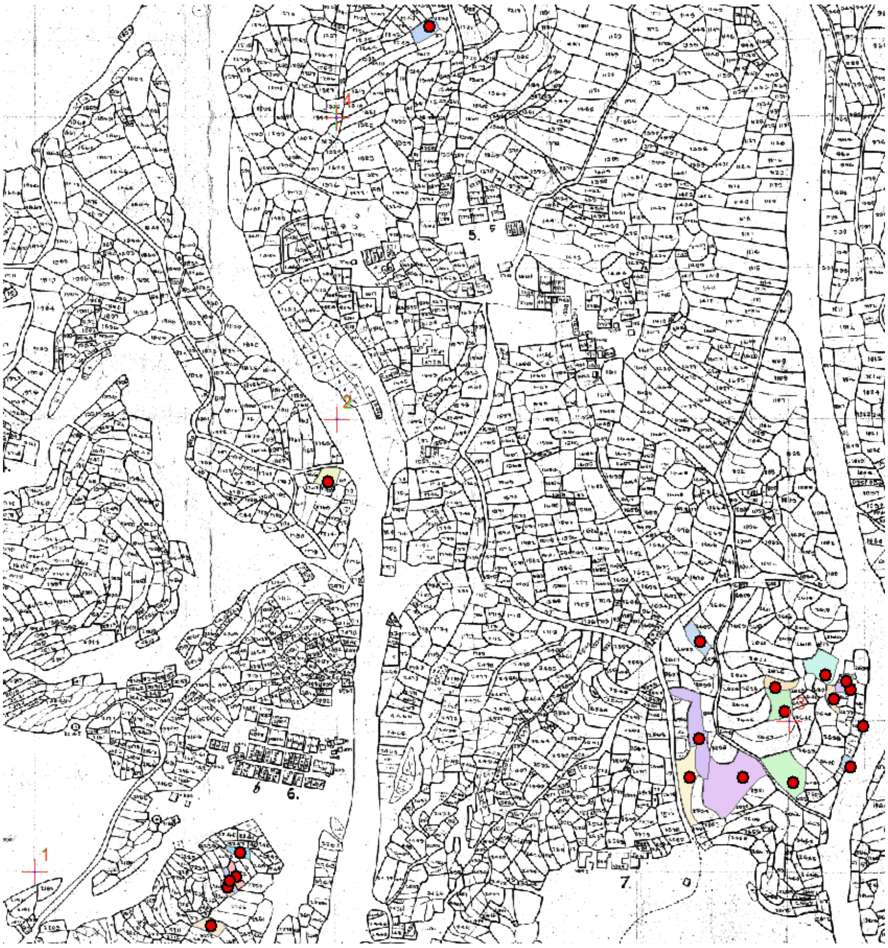
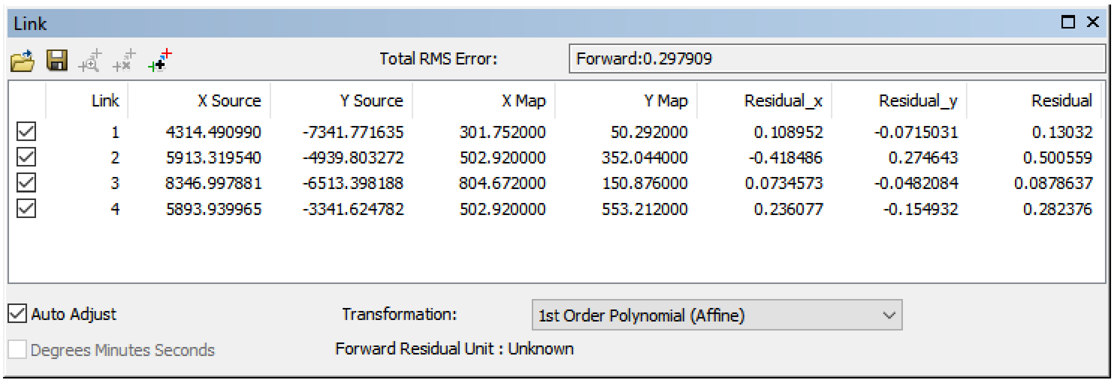


Figure 4-2: Geo-referencing of DD 1 TC



Link	X Source	Y Source	X Map	Y Map	Residual_x	Residual_y	Residual
1	4314.490990	-7341.771635	301.752000	50.292000	0.108952	-0.0715031	0.13032
2	5913.319540	-4939.803272	502.920000	352.044000	-0.418486	0.274643	0.500559
3	8346.997881	-6513.398188	804.672000	150.876000	0.0734573	-0.0482084	0.0878637
4	5893.939965	-3341.624782	502.920000	553.212000	0.236077	-0.154932	0.282376

Figure 4-3: The link table

4.2.2 DD 129

DD 129 was plotted at 16 inches to 1 mile, which is 1:3960. A gridline the DD sheet has a length of 1 inch. To turn the length of a gridline into actual scale, in meter unit, the following approaches were conducted: -

1. 16 inches to 1 mile = 16 inch to 1609.344 m

2. 16 inches to 1609.344 m → 1 inch to 100.584 m

The gridline in a 1:1980 DD sheet equals to 50.292 m. Figure 4-2 shows the geo-referencing of DD 1 TC. The points in red are the selected 20 lots. The red cross are the control points. The total Root Mean Square (RMS) error was 0.111173 (Figure *4-3*)

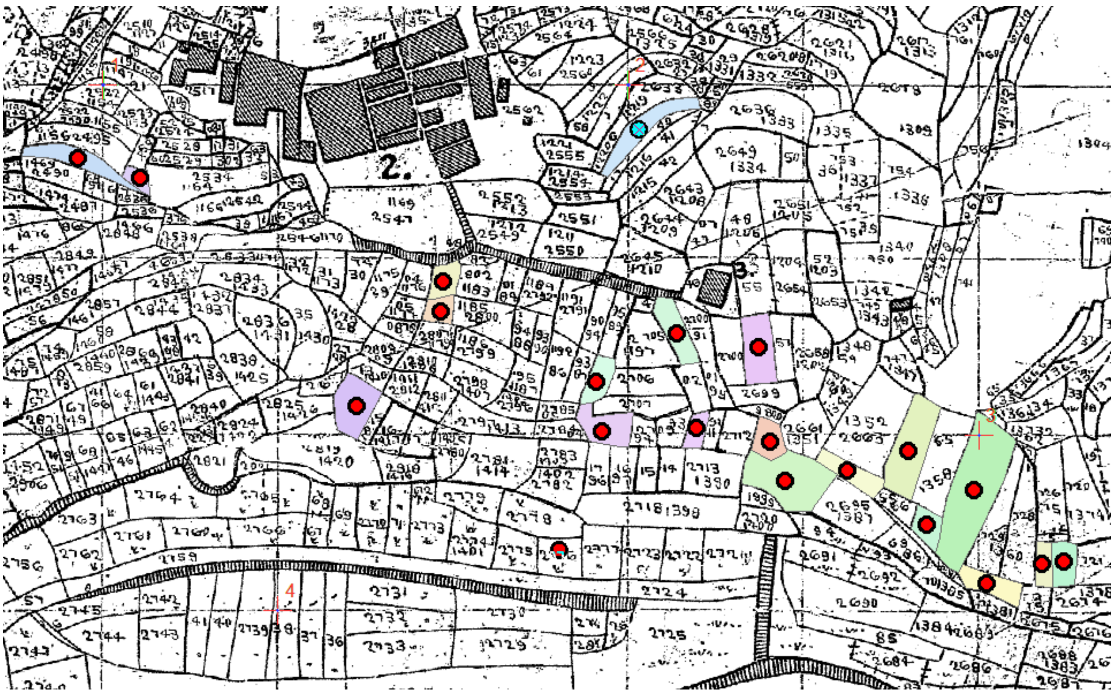
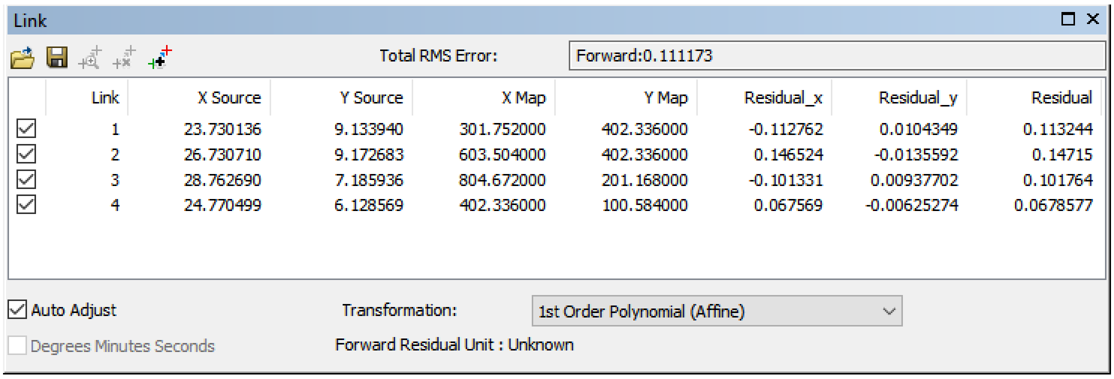


Figure 4-4: Geo-referencing of DD 129



Link

Total RMS Error: Forward:0.111173

	Link	X Source	Y Source	X Map	Y Map	Residual_x	Residual_y	Residual
☑	1	23.730136	9.133940	301.752000	402.336000	-0.112762	0.0104349	0.113244
☑	2	26.730710	9.172683	603.504000	402.336000	0.146524	-0.0135592	0.14715
☑	3	28.762690	7.185936	804.672000	201.168000	-0.101331	0.00937702	0.101764
☑	4	24.770499	6.128569	402.336000	100.584000	0.067569	-0.00625274	0.0678577

☑ Auto Adjust Transformation: 1st Order Polynomial (Affine)

☐ Degrees Minutes Seconds Forward Residual Unit : Unknown

Figure 4-5: Link Table

4.3 Lot Area and Dice Similarity Coefficient

As mentioned in Section 2.2, one of the procedures in the lot boundary re-establishment is to compare the surveyed area with the register area of the lot. It aims to ensure the surveyed area is consistent with the register area, in terms of dimension. The allowable tolerance is: -

For DD sheets of 1:1980, the discrepancy is suggested to be within:

+/- (20 sq. metres + 5% of the registered area)

For DD sheets of 1:3960, the discrepancy is suggested to be within:

+/- (20 sq. metres + 10% of the registered area)

The register area can be found in Schedule book (Figure 4-6). After the comparison, the lots will be classified into two types: lot within tolerance and lot not within tolerance. Then, the DSC of those two types of lots will be computed. If the lot is within tolerance but have a low DSC, it means that the lot's DD's area is close to its surveyed area but not similar in terms of shape. If the lot is not within tolerance but have a high DSC, it means that the lot's DD's area is not close to its surveyed area but similar in terms of shape.

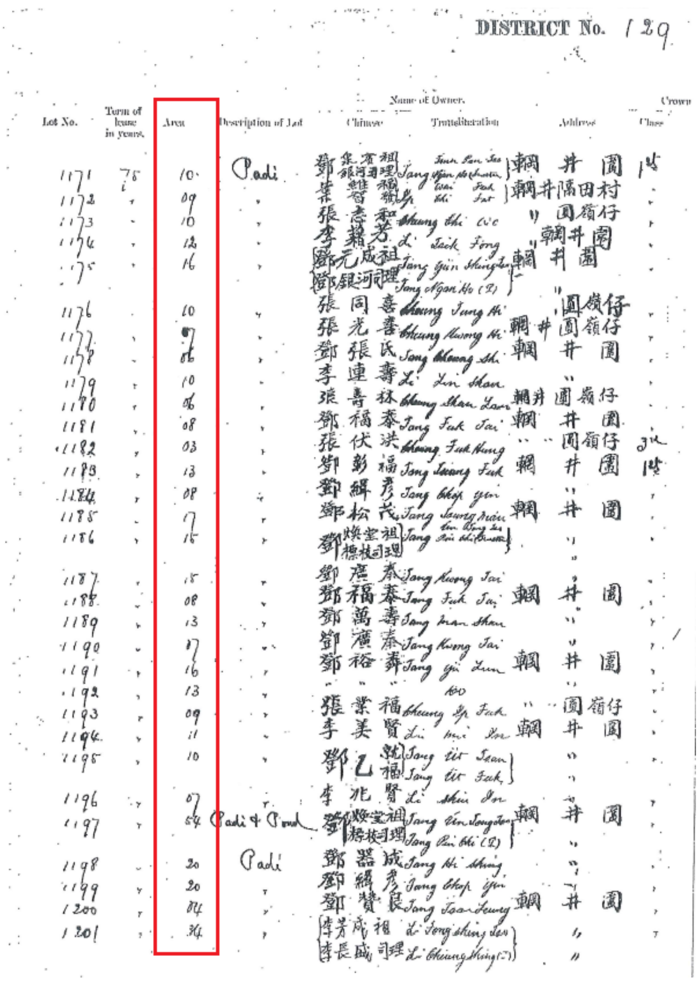


Figure 4-6: Schedule Book of Lot 1171 - 1201

4.3.1 DD 1 TC

Table 4-2 shows the comparison between the register area and surveyed area of the 20 lots in DD 1 TC. Table 4-3 shows the DSC of these 20 lots. A summary of the result is presented below: -

1. 17 of the lots are within tolerance while 3 of the lots are not within tolerance;
2. Lot 2636 has a DSC of 0.98, which is the highest. lot 2651 a DSC of 0.97, which is the second-highest. Their areas are within the allowable tolerance;
3. Lot 2237 has a DSC of 0.76, which is the lowest. Its area is within the allowable tolerance;
4. The area of 3 lots are not within allowable tolerance but have a DSC above 0.91. They are lot 2248, 2255, 2627;
5. 19 of the lots have a DSC above 0.89. The average is 0.92.

Table 4-2: Area Comparison of 20 selected lots in DD 1 TC

Lot	Register Area (acre)	Register Area (sq. m)	Surveyed Area (sq. m)	Discrepancy (SA – RA) (sq. m)	Allowable Tolerance (sq. m) (+/-)	Within Tolerance
1311	0.05	202.3	180.1	-22.2	30.1	Yes
1765	0.04	161.9	141.5	-20.4	28.1	Yes
2237	0.01	40.5	45.2	4.7	22.0	Yes
2248	0.02	80.9	110.0	29.0	24.0	No
2254	0.02	80.9	89.7	8.8	24.0	Yes
2255	0.01	40.5	75.7	35.2	22.0	No
2257	0.04	161.9	158.5	-3.3	28.1	Yes
2593	0.25	1011.7	987.5	-24.2	70.6	Yes
2595	0.10	404.7	414.6	9.9	40.2	Yes
2597	0.15	607.0	580.4	-26.6	50.4	Yes
2603	0.04	161.9	181.4	19.5	28.1	Yes
2627	0.04	161.9	108.9	-53.0	28.1	No
2629	0.08	323.7	334.8	11.0	36.2	Yes
2631	0.02	80.9	84.8	3.9	24.0	Yes
2632	0.02	80.9	69.4	-11.5	24.0	Yes
2633	0.01	40.5	40.5	0.1	22.0	Yes
2636	0.03	121.4	115.3	-6.1	26.1	Yes
2643	0.07	283.3	261.5	-21.7	34.2	Yes
2648	0.13	526.1	557.1	31.0	46.3	Yes
2651	0.01	40.5	46.6	6.1	22.0	Yes

Table 4-3: DSC of the 20 lots in DD 1 TC

Lot	Within Tolerance	Area in DD Sheet (sq. m)	Area in SRP (sq. m)	Overlapped Area (sq. m)	DSC
1311	Yes	178.3	180.1	172.2	0.96
1765	Yes	138.7	141.5	133.8	0.96
2237	Yes	29.8	45.2	28.4	0.76
2248	No	103.4	110.0	99.2	0.93
2254	Yes	86.3	89.7	82.4	0.94
2255	No	69.6	76.5	66.8	0.91
2257	Yes	148.1	158.5	141.4	0.92
2593	Yes	1004.6	987.5	929.6	0.93
2595	Yes	395.8	414.6	368.9	0.91
2597	Yes	568.8	580.4	518.7	0.90
2603	Yes	170.9	181.4	155.8	0.88
2627	No	106.8	108.9	98.7	0.92
2629	Yes	325.5	334.8	314.7	0.95
2631	Yes	73.3	84.8	68.7	0.87
2632	Yes	66.5	69.4	62.7	0.92
2633	Yes	38.4	40.5	36.6	0.93
2636	Yes	114.8	115.3	112.3	0.98
2643	Yes	256.9	261.5	245.9	0.95
2648	Yes	547.4	557.1	518.9	0.94
2651	Yes	45.3	46.6	44.4	0.97

Lot 2636 and Lot 2651

Lot 2636 has a DSC of 0.98, which is very close to 1. It means that the boundary depicted on DD sheet and the boundary on the Survey Record Plan are very similar to each other. The result is shown in Figure 4-7 (a). The survey report of lot 2636 states that the boundaries were defined in accordance with the position of DD in the absence of the original boundary evidence. Therefore, it is reasonable that a high DSC was computed.

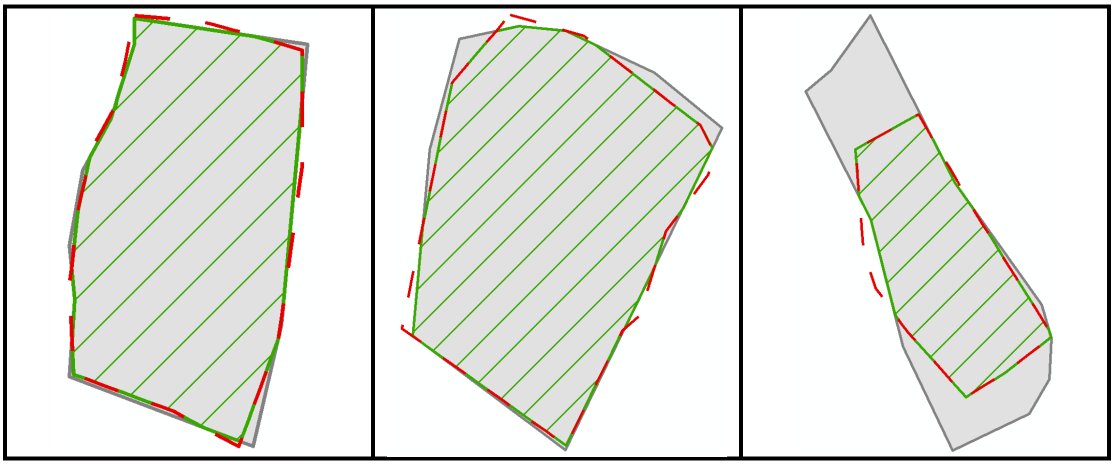
Lot 2651 has a DSC of 0.97, which is also very close to 1. The result is shown in Figure 4-7 (b). The respecting survey report stated that 3 boundary points defined generally along the field bund as shown on old aerial photo consistent with the DD sheet. The remaining boundaries were defined in accordance with the position on the DD sheet, in the absence of any reliable boundary evidence. Like the Lot 2636, it is reasonable that Lot 2651 has the second-highest DSC among the others because their boundaries were defined in accordance with the position on the DD sheet. It proved that DSC can reflect the true information.

Lot 2237

Lot 2237 has a DSC of 0.79, which is the lowest among the others. It means that the boundary depicted on DD sheet and the boundary on SRP are not much similar to each other. The result is shown in Figure 4-7 (c). The DD sheet's boundary is much smaller than the SRP's boundary. The survey report of lot 2237 states that: -

> "The field bunds observed from the year 1963 orthophoto, which tally with the boundary depiction on DD sheet, are regarded as boundary features and are adopted as the subject lot boundary in general."
>
> "The existing rubble walls found along the north-eastern and south-western bounds of the lot, which tally with the field bund shown on the year 1963 orthophoto, are regarded as existing boundary features. The bottom of the rubble walls are adopted as C-E and E-A sides of the boundary."

From Figure 4-7 (c), it is obvious that the northern and southern position of the boundary in SRP is far away from the position depicted on the DD sheet. According to the survey report, the northern and southern side of the boundary was defined in accordance with the rubble walls on the field bunds. Some questions can be raised for this situation: Will the field bund or rubble walls change after the time of grant? Most of the old schedule lots are granted in 1905. It has been around 58 years, from 1905 – 1963. The other reason may be the error of the DD sheet. The message can be presented here is low DSC can act as a referencing index. True, the lot area comparison is within the allowable tolerance, but the DSC is comparatively lower than the others. DSC can remind the land surveyors to further consider the lot boundary determination.



(a) Lot 2636 (b) Lot 2651 (c) Lot 2237

Figure 4-7: Result from the shape context descriptor: the grey polygon represents the boundary shown in SRP, the dashed line represents the boundaries from the DD sheet after 2D affine transformation and the green hatched area represent the overlapping area.

Lot 2248 and Lot 2255

Lot 2248, 2255 are the lots which their surveyed areas are not within the allowable tolerance. The correlation result is shown in Figure 4-8. Lot 2248 and Lot 2255 have a DSC of 0.93 and 0.91 respectively. The survey report of Lot 2248 stated: -

> "The field bunds observed from the year 1963 orthophoto, which resemble with the boundary depiction on D.D. sheet, are regarded as boundary features and are adopted as the subject lot boundary in general.
>
> The existing rubble walls found along the north-eastern bound of the lot, which tally

with the field bund shown on the year 1963 orthophoto, are regarded as existing boundary features. The top of the rubble wall is adopted as N-S side of the boundary."

The survey report of Lot 2255 stated: -

> "The field bunds observed from the year 1963 orthophoto, which tally with the boundary depiction on D.D. sheet, are regarded as boundary features and are adopted as the subject lot boundary in general.
>
> The existing rubble walls found along the north-eastern and south-western bounds of the lot, which tally with the field bund shown on the year 1963 orthophoto, are regarded as existing boundary features. The bottom of the rubble walls are adopted as Y-AJ and AJ-AS sides of the boundary."

The boundaries of lot 2248 and 2255 were defined in accordance with the field bund and rubble wall. The surveyed area of lot 2248 is 28 sq. m larger than the register area, while the surveyed area of lot 2255 is 35.2 sq. m larger than the register area. Although their area not within tolerance, the DSC are comparatively higher. It means that the two boundaries are consistent in terms of shape. DSC can act as a referencing index to support the interpretation of a land surveyor, whether the re-established boundary is appropriate or not.

Lot 2627

Lot 2627 has a DSC of 0.91. The respecting survey report mentioned: -

> "Boundary lines FE-EQ-ER were fixed by adpoting the corresponding common boundary data of adjoining lot with previous survey record. The remaining boundaries were defined by following the field bund as shown on old aerial photo no. 4640 consistent with the DD.
>
> There might be some scale errors in the area on the BCL, which resulted that the surveyed area of Lot 2627 is significantly smaller than the registered area by 0.013 ac."

The boundary was established by following the previous survey record and the field bund. Similarly, the DSC can support the land surveyors on the boundary determination, even though the area is smaller than the registered area by 53 sq. m.

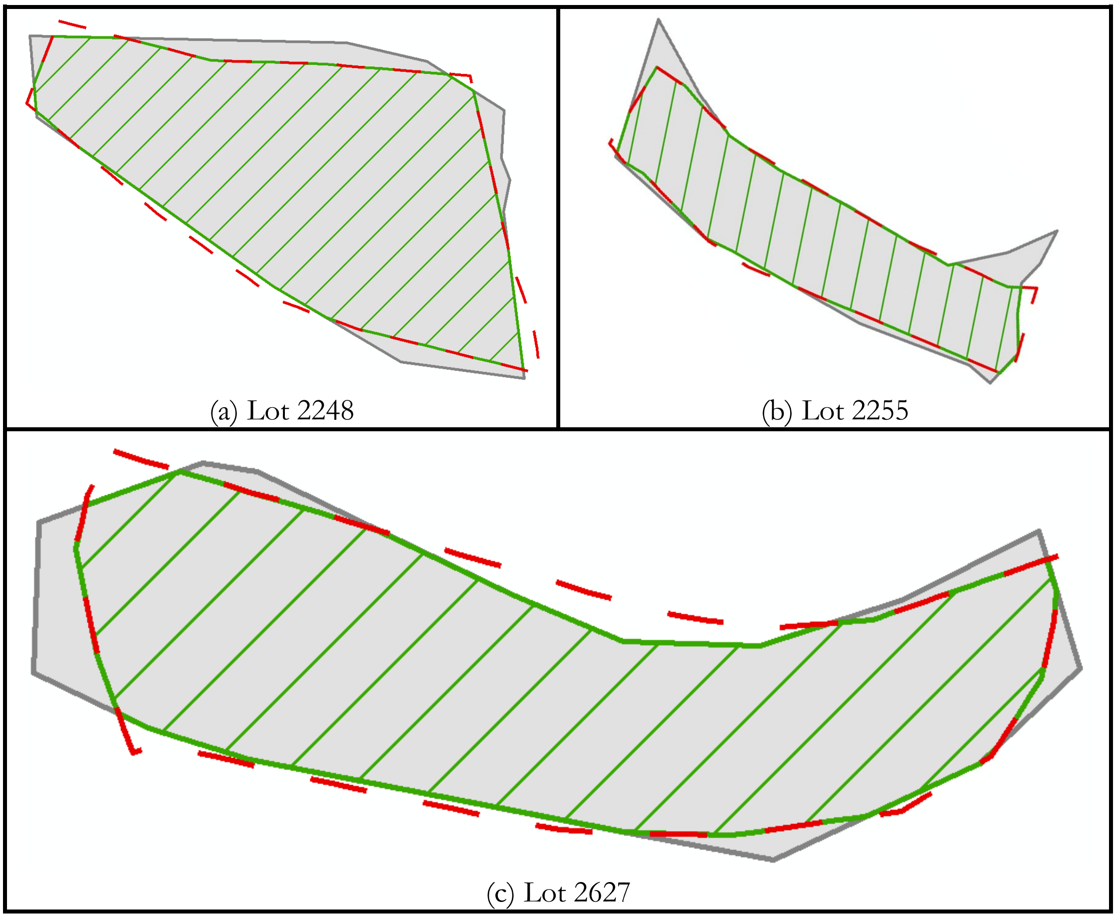


Figure 4-8: Result from the shape context descriptor: the grey polygon represents the boundary shown in SRP, the dashed line represents the boundaries from the DD sheet after 2D affine transformation and the green hatched area represent the overlapping area.

4.3.2 DD 129

Table 4-4 shows the result of comparison between register area and surveyed area of the 20 lots in DD 129. shows the DSC of the 20 lots. Table 4-5 shows the DSC of those 20 lots. Some points can be highlighted: -

1. 16 of the lots are within tolerance while 4 of the lots are not within tolerance;
2. 7 of the lots have a DSC of 0.96, which are the highest among the others;
3. 3 of the lots have a DSC of 0.91, which are the lowest;
4. The 4 lots which are not within allowable tolerance have a DSC above 0.93;
5. All lots have a DSC above 0.91. The average is 0.94;

Table 4-4: Area comparison of 20 selected lots in DD 129

Lot	Register Area (acre)	Register Area (sq. m)	Surveyed Area (sq. m)	Discrepancy (SA - RA) (sq. m)	Allowable Tolerance (+/-)	Within Tolerance
1158	0.04	161.9	149.8	-12.0	36.2	Yes
1180	0.06	242.8	319.3	76.5	44.3	No
1181	0.08	323.7	316.0	-7.8	52.4	Yes
1196	0.07	283.3	293.8	10.5	48.3	Yes
1198	0.20	809.4	804.6	-4.8	100.9	Yes
1201	0.34	1375.9	1764.4	388.4	157.6	No
1218	0.14	566.6	551.3	-15.3	76.7	Yes
1353	0.10	404.7	378.5	-26.2	60.5	Yes
1354	0.25	1011.7	1063.7	52.0	121.2	Yes
1357	0.08	323.7	303.2	-20.6	52.4	Yes
1359	0.49	1983.0	1972.0	-11.0	218.3	No
1376	0.06	242.8	264.2	21.4	44.3	Yes
1377	0.08	323.7	356.7	33.0	52.4	Yes
1380	0.10	404.7	454.9	50.2	60.5	Yes
1388	0.28	1133.1	1039.9	-93.2	133.3	Yes
1389	0.09	364.2	360.3	-3.9	56.4	Yes
1392	0.05	202.3	286.8	84.5	40.2	No
1395	0.12	485.6	535.6	50.0	68.6	Yes
1399	0.05	202.3	241.7	39.4	40.2	Yes
1419	0.17	688.0	739.4	51.4	88.8	Yes

Table 4-5: DSC of the 20 selected lots in DD 129

Lot	Within Tolerance	Plan Area (sq. m)	Surveyed Area (sq. m)	Overlapped Area (sq. m)	DSC
1158	Yes	146.0	149.8	134.0	0.91
1180	No	298.8	319.3	288.1	0.93
1181	Yes	297.5	316.0	286.1	0.93
1196	Yes	287.6	293.8	278.3	0.96
1198	Yes	782.2	804.6	758.1	0.96
1201	No	1780.7	1764.4	1701.0	0.96
1218	Yes	578.7	551.3	512.3	0.91
1353	Yes	367.5	378.5	359.0	0.96
1354	Yes	1039.4	1063.7	1005.9	0.96
1357	Yes	299.6	303.2	284.2	0.94
1359	No	1912.0	1972.0	1850.7	0.95
1376	Yes	251.8	264.2	244.1	0.95
1377	Yes	346.2	356.7	336.8	0.96
1380	Yes	453.9	454.9	424.8	0.93
1388	Yes	982.9	1039.9	955.9	0.95
1389	Yes	348.4	360.3	338.2	0.95
1392	No	281.7	286.8	272.2	0.96
1395	Yes	505.1	535.6	477.1	0.92
1399	Yes	220.9	241.7	211.3	0.91
1419	Yes	725.4	739.4	696.8	0.95

Lot 1196 and Lot 1198

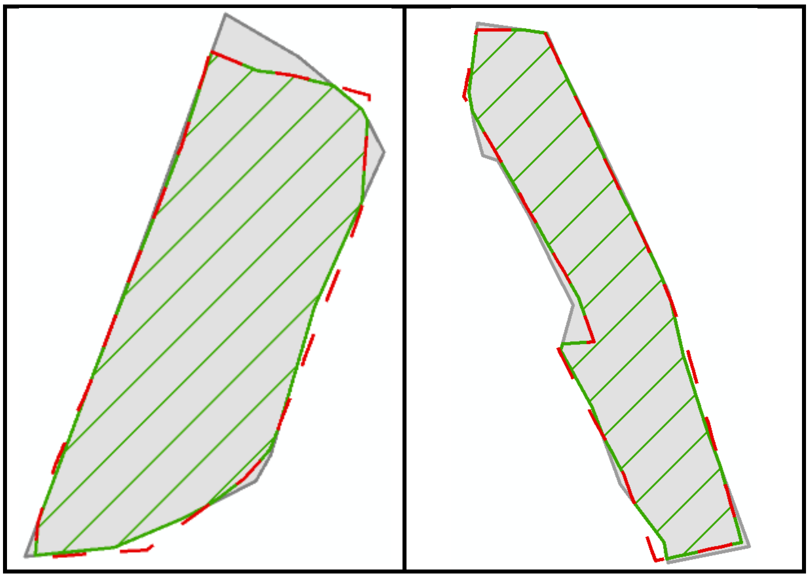
Lot 1196 has a DSC of 0.96. Its survey report stated that: -

> "After evaluation of the available evidence, boundary lines C-D-E-F-G-H and K-L-A were fixed by adopting the corresponding common boundary data of adjoining lots with previous survey records. The remaining boundaries A-B-C and H-J-K of the subject lot were defined generally along the field bunds as shown on old aerial photo of 1945 of old boundary evidence"

Lot 1198 also has a DSC of 0.96. Its survey report stated that: -

> "The final defined boundaries of the subject lot corresponding generally with the field bunds as shown on old aerial photo no. 3105 of 1945 and the lot index plan"

The boundary of Lot 1196 was defined in accordance with the previous survey record and the field bunds, while the boundary of Lot 1198 was defined in accordance with field bund only. As shown in Figure 4-9, the boundaries are very similar to each other and highly overlapped together. As mentioned in Section 2.1, the accuracy of DD sheet is low, not to mentioned it is plotted in 1: 3960. It is just an approximate location of the field bunds. In these two lot cases, DSC can act as a supporting index and give the land surveyor confidence for his or her boundary determination.



(a) Lot 1196 (b) Lot 1198

Figure 4-9: Result from the shape context descriptor: the grey polygon represents the boundary shown in SRP, the dashed line represents the boundaries from the DD sheet after 2D affine transformation and the green hatched area represent the overlapping area.

Lot 1180 and 1201

Lot 1180 has a DSC of 0.93. It is the lowest among the 4 lots which are not within allowable tolerance. Its survey report mentioned: -

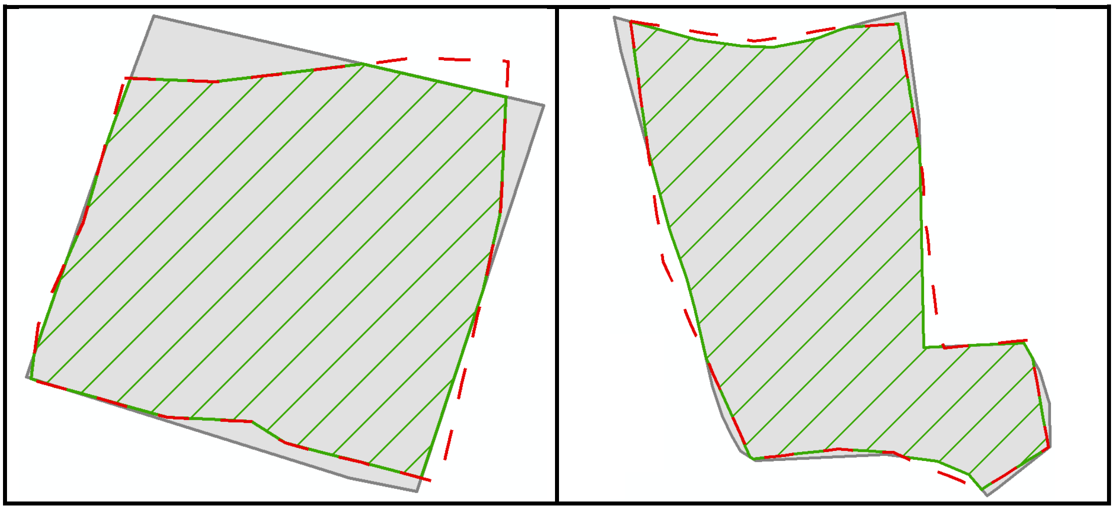
> "The surveyed features were also compared on an overlay. It was noted that the southern and western walls of Lot 1180 as well as the concrete footpath at the eastern side of Lot 1180 were general matched with the corresponding boundary field bunds as identified in DOP 1000-1963. In this connection, the above existing occupation was believed to be established on the old field bunds and could be regarded as boundary features for boundary re-establishment."

Lot 1180 was defined in accordance with the existing occupation: footpath and wall. However, the surveyed area is larger than the register area by 76.5 sq. m. There is a possibility that the footpath and wall were re-built after the time of grant (i.e. during 1905 – 1963), which means that the existing occupation may not be the occupation at the time of grant. Nevertheless, it is impossible to find the occupation at the time of grant since no aerial photographs at 1905 were available. By using shape context descriptor, 0.93 of DSC is computed, which is close to 1. It could support the decision of land surveyors on re-establishing boundary, to a certain extent. Although the area is not within allowable tolerance, the re-established boundary is still acceptable.

Lot 1201 has a DSC of 0.96. It is the highest among the 4 lots which are not within allowable tolerance. Its survey report stated: -

> "With the defined boundaries which generally agreed with the aforesaid records, the common boundaries with adjoining Lot Nos 1200 … were adopted. The northern boundaries were tallied with the edge of existing footpath while the western boundaries were along the mid-way of footpath as matched with the correlated boundaries of the said Cadastral Sheets. Lastly, the remaining boundaries were re-established so as to agree with the shape of the DD delineation"

The surveyed area of Lot 1201 is 388.4 sq. m larger than the registered area, which is not within the allowable tolerance. Then, the shape of the boundary can be checked by the shape context descriptor. It is observed that the surveyed area is very similar to the area depicted in DD sheet. (Figure 4-10 b). Also, a high DSC – 0.96, was computed. The surveyed area was highly consistent with the DD sheet's area, which is very important. As mentioned in Section 2.2, the Indenture attached in the Block Crown Lease stated that a piece of land parcel was "more particularly delineated and described on the plan attached. The evidence in DD sheet is more important than the register area. In this case, shape context descriptor acts as a strong supporting evidence to the re-established boundary.



(a) Lot 1180 (b) Lot 1201

Figure 4-10: Result from the shape context descriptor: the grey polygon represents the boundary shown in SRP, the dashed line represents the boundaries from the DD sheet after 2D affine transformation and the green hatched area represent the overlapping area.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Lot boundary re-establishment is an important work in the Hong Kon Cadastral System. It is highly demanded because of the primary defects of the land grant documents. Area discrepancy checking may not be enough to support a decision made by the land surveyors on the land boundary determination. Shape context descriptor can also be adapted. The computed Dice Similarity Coefficient (DSC) can act as a reference, as the area discrepancy checking. After using 40 lots in the experiment, the conclusion is made below: -

1. **Shape context descriptor can minimise human error.**
   In the current practice, the correlation works are fully done by a human. For the procedure of comparison DD sheet and Survey Record Plan. Firstly, the boundary depicted on DD sheet and survey area will be digitised. Secondly, the two boundaries will be correlated with each other. These two steps are operated by a human, which may involve human error. If shape context descriptor is performed, the correlation will be done by geo-registration techniques like 2D Affine Transformation. It minimises human error.

2. **Similar performance of the shape context descriptor on DD sheets plotted in 1:1980 and 1:3960 was found.**
   It is observed that the DSC in DD sheet plotted in 1:1980 and the DSC in DD sheet plotted in 1:3960 are close to each other. The average of DSC in DD sheet plotted in 1:1980 is 0.92. The average of DSC in DD sheet plotted in 1:3960 is 0.94, which is slightly higher than that in DD sheet plotted in 1:1980. It may be due to the detail in DD sheet plotted in 1:3960 is comparatively rough. Therefore, the shape of the boundary shows a simpler geometry, which is easier to match with the respecting surveyed area.

3. **DSC can become a standard to check the accuracy of defined boundary**
   It is well known that a boundary determination should not strictly follow either the register area or the area depicted in DD sheet. There is a lot of other evidence such as the aerial photograph, old survey sheet, existing occupation etc. However, the register area and the area depicted in DD sheet are still used as a reference. They are the product of the land grant documents. This is the reason that discrepancy checking on area is necessary. As mentioned in Section 2.2, this standard has a defect because it only focusses on the area but not shape. Two polygons can have the same area but one of them is a rectangle and the other one is a pentagon. DSC can provide a reference regarding the shape of the boundaries.

5.2 Future Work

It still takes a long time for applying the shape context descriptor into lot boundary determination. The whole algorithm is not mature enough. For example, the user interface is inadequate. If someone does not know Python programming language, he or she will not know how to conduct the algorithm. Therefore, a user interface must be developed. Also, the computation procedure is too clumsy. The program script requires a lot of effort in the aspect of input and process. The structure of the script should be modified to attain a more efficient workflow.

In this dissertation, the main purpose is to create a shape context descriptor for the use of boundaries correlation and test the feasibility of using the descriptor in the current land boundary survey system in Hong Kong. Further investigation can be done in the future. For example, a value v should be found. The DSC lower than v will be defined as not within tolerance while the DSC higher than v is defined as within tolerance. In this stage, more sample size, even all of the lots depicted on one DD sheet, should be used in the experiment.

The shape context descriptor was used to compare the area depicted in DD sheet with the surveyed area. It can also be used for the comparison between any other documents such as DD sheet and aerial photograph, surveyed area and aerial photograph etc. The algorithm is very flexible. It is hoped that more computer vision algorithm will be developed to assist the land boundary survey. “Smart City” is the theme that the government focus on, in this century. Common Spatial Data Infrastructure will be a key component to transform Hong Kong into a “Smart City”. A million of land and survey record will be transformed into another platform, which might need help from the computer vision. The computer vision algorithms like shape context, must be the necessary tools for future development.

Reference

[1] Alterovitz, R., Goldberg, K., Pouliot, J., Hsu, I., Kim, Y., Noworolski, S., & Kurhanewicz, J. (2006). *Registration of MR prostate images with biomechanical modeling and nonlinear parameter estimation*.

[2] Anupoju, S. (2019, March 4). Errors in Plane Table Surveying. Retrieved from https://theconstructor.org/surveying/plane-table-surveying-errors/30430/

[3] Belongie, & Malik. (2000). Matching with shape contexts. *2000 Proceedings Workshop on Content-based Access of Image and Video Libraries*, 20-26.

[4] Belongie, S., Mori, G., Malik, J., Krim, H., & Yezzi, A. (2006). Matching with Shape Contexts. In *Statistics and Analysis of Shapes* (Modeling and Simulation in Science, Engineering and Technology, pp. 81-105). Boston, MA: Birkhäuser Boston.

[5] Couling, S. (1917). *Encyclopaedia sinica*. Shanghai: Kelley and Walsh.

[6] Dice, L. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3), 297-302.

[7] Hong Kong Government, (1902d), Appendix No.2, Report on the Survey of the New Territory, At the close of the Field Season of 1900-01-15th July, 1901 by Geo. P. Tate, *The Hong Kong Government Gazette*, 2nd May, 1902.

[8] Lam Pak Kau v. Yu Yuet Fat. [1996]. N. A2020.

[9] Lands Department. (2006). *Code of Practice Land Survey Ordinance (Chapter 473)*. Hong Kong: Lands Department.

[10] Lands Department. (2019). *District Survey Office Technical Manual*. Hong Kong: Lands Department.

[11] Newland, W. J. (1904), A General Report of the Survey of the New Territories from November 1899 to April 1904, *Hong Kong Government Gazette*, 1904.

[12] Oliveira, F., & Tavares, J. (2014). Medical image registration: A review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2), 73-93.

[13] Paul B. (2014). *GIS Fundamentals: A First Text on Geographic Information Systems, NEW*

and UPDATED Sixth Edition (4th ed.). Minnesota. University of Minnesota.

[14] Yan, W., Easa, S., & Shaker, A. (2017). Polygon-based image registration: A new approach for geo-referencing historical maps. *Remote Sensing Letters*, 8(7), 703-712.