AN IMPROVED WATERSHED TRANSFORM ALGORITHM FOR TWO-HAND TRACKING UNDER PARTIAL OCCLUSION

LEIM PENG PENG

UNIVERSITI SAINS MALAYSIA

2016

AN IMPROVED WATERSHED TRANSFORM ALGORITHM FOR TWO-HAND TRACKING UNDER PARTIAL OCCLUSION

by

LEIM PENG PENG

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

May 2016

ACKNOWLEDGEMENT

I would like to express my gratitude to my supervisor, Pn. Tan Guat Yew for her generous support and excellent mentorship throughout the duration of this project. Also to thank Dean, lecturer and staff of School of Mathematical Sciences and Institute of Postgraduate for providing me a comfortable place for study and their generous help in various ways for the completion of my study.

Moreover, I am grateful to the grant 304/PMATHS/6312096 and mybrain for providing me with the financial support to undergo my post graduate Master degree at the University of Science, Malaysia. Lastly, my sincere thanks go to my beloved parents, siblings and friends for their endless support and motivation in the pursuit of my dreams. Thank you everyone.

TABLE OF CONTENTS

Page

LIST OF FIGURES

PENAMBAHBAIKAN ALGORITMA TRANSFORMASI TADAHAN AIR BAGI PENJEJAKAN DUA-TANGAN DI BAWAH OKLUSI SEPARA

ABSTRAK

Untuk mencapai interaksi semulajadi dalam persekitaran realiti penambahan, berinteraksi dengan isyarat dua-tangan selalunya dijadikan pilihan yang utama. Namun, interaksi dua-tangan tersebut akan mengakibatkan kedua-dua tangan saling beroklusi dan mengganggu penjejakan isyarat tangan. Dalam kajian ini, kami telah mencadangkan suatu penyelesaian untuk menjejak isyarat-isyarat tangan yang saling beroklusi dengan menambahbaikkan algoritma transformasi tadahan air dan menjana sistem penjejakan dua-tangan lanjutan berdasarkan sistem pengenalian dasar-penglihatan. Prosedur untuk menyelesaikan pertindihan dua tangan bermula dengan pengesanan warna kulit bagi memperoleh rantau minat yang kemudiannya dianggap sebagai calon rantau tangan. Ini diikuti dengan pengiraan kecerunan imej dari satu imej berwarna kelabu untuk dapat sempadan dua tangan. Seterusnya, satu titik benih diekstrak daripada imej input untuk proses banjir. Proses banjir bermula dari titik benih tersebut and berakhir apabila ia sampai sempadan. Pixel yang diisi dalam proses banjir adalah dianggap sebagai rantau daripada salah satu tangan. Kaedah penolakan digunakan untuk mengekstrak rantau tangan kedua. Untuk pembangunan system pengesanan tangan, empat senario pengesanan tangan yang berkemungkinan dipertimbangkan: i) tiada tangan, ii) satu tangan, iii) dua tangan yang berasing, dan iv) dua tangan yang bertindih. Dalam proses pengesanan, kami juga melaksanakan pengekstrakan ciri-ciri tangan bagi memastikan kedua-dua rantau output ada rantau tangan.

AN IMPROVED WATERSHED TRANSFORM ALGORITHM FOR TWO-HAND TRACKING UNDER PARTIAL OCCLUSION

ABSTRACT

To achieve a natural interaction in augmented reality environment, two-handed gesture interactions are highly preferred. However, two-handed interactions always result in mutual occlusions which interfere with the hand gesture recognition. In this research, a solution for this problem is presented by improving the watershed transform algorithm and developing an advanced two-hand tracking system based on vision-based recognition system. The procedure of solving two overlapping hands starts with skin colour detection to acquire the regions of interest which are then assumed to be the candidates of the hands' regions, followed by the computation of image gradients from a grey colour image to obtain the boundaries of two hands. Next, a seed point is extracted from input image for flooding process. Flooding process is started from the seed point and it ends when it reaches the boundary. The pixels filled in the flooding process are considered as the region of one of the hands. Subtraction method is then applied to extract the second hand region. For hand tracking system development, four possible scenarios of hand tracking are considered: i) no hand, ii) one hand, iii) two separate hands, and iv) two overlapping hands. In the tracking process, we also implemented hand feature extraction to ensure that both output regions are hand region.

CHAPTER 1

INTRODUCTION

1.1 Background Study

Hand is one of the most important parts of a human body that allows him/her to carry out daily activities which include communications. As such, hand gesture recognition researches have gained prominence in Human-Computer Interaction (HCI) over the recent years because of its extensive applications in human-machine interface systems and virtual environments.

HCI is a field of research on how humans interact with computers through a user interface which comprises both software and hardware. For instance, software is used to display input features on computer's monitor while hardware such as keyboard and mouse allows users to insert input. Although the invention of keyboard and mouse is a great progress in HCI, these input hardware only support 2-dimensional (2D) interactions; it may not be suitable for natural 3-dimensional (3D) interactions in virtual environment applications. Multidimensional input devices such as joystick and 3D mouse are invented for 3D interactions, but these devices do not mimic the naturalness in human-computer interaction as users do not interact with virtual objects directly. Due to this reason, research of using hand gestures as input devices is becoming popular as it is a more intuitive method in HCI. By identifying 3D hand positions and recognizing specific hand gestures, it is able to enhance selection, manipulation, and navigational tasks to generate information in communicating with the computer. Some of the methods to track this information are sensor-based and

vision-based. Sensor-based method, although high in detection accuracy, involves expensive devices to detect hand gestures. Also, infrared devices and data gloves that are usually required limit the freedom of hand movements due to their attachments to the fingers. By contrast, the vision-based method involves recording of the bare hand movements with video cameras, of which the hand gestures in the images are recognized and analysed with image processing algorithms. Although vision-based method enables bare hand recognition without the need of expensive and restrictive devices in sensor-based method, the occlusion problem still remains. On the projected 2D image plane, two objects that are spatially separated in 3D space may occlude each other and cause two-handed gestures are unable to be recognized. To distinct an image into meaningful structures, image segmentation often plays an important role in solving two-hand occlusion problems.

Image segmentation is a process that partitions an image into its constituent parts and extracts those parts of interest or objects, such that each region of interest is homogeneous with respect to some property, such as colour, intensity, or texture. A lot of segmentation methods had been studied previously, and categorization is necessary to outline the methods properly. Most of the earlier literature had categorized the segmentation methods as below [1]:

a. Pixel-based segmentation is the method of dividing a digital image into multiple segments or groups of pixels.
b. Threshold-based segmentation. This segmentation method is a combination of histogram thresholding and slicing techniques. This method may be applied on an image directly; either combined with pre-processing or post-processing techniques.

c. Edge-based segmentation. Edges detected in an image are assumed to represent object boundaries which are used for object recognition.
d. Region-based segmentation. This segmentation method examines neighbouring pixels of initial seed points and decides whether these pixels should be added to the region. The regions are then spread from these seed points to adjacent points based on a region membership criterion. The criterion could be pixel intensity, greyscale texture, or colour.

In order to segment the homogenous colour distribution of both hands, a hybrid technique known as morphological watershed transform is proposed and studied. The watershed transform is a combination of both edge-based and region-based segmentation approaches [2]. The fundamental of watershed transform is visualizing a grey level image into its topographic representation, where the height of each point is corresponding to the intensity of grey level. The greater the pixel values, the higher gradient of ridge area in image; low gradient pixels are considered as basin. The idea of watershed algorithm is like a water stream falling from the ridge. It flows along a path to reach the catchment basin to form a region. The watershed lines divide the catchment sink into different areas which form the object contours in the image. Although watershed transform has been frequently used in current fields of image processing, however, the main issue of its sensitivity to noise that causes over-segmentation still remains [3].

The main objective of this thesis is to illustrate the occlusion problem we have faced and how we solve it by improving the watershed transform algorithm in the case of two-handed interaction.

1.2 Motivation

Interactive Augmented Reality (AR) possess the ability to display, interact and control of the virtual 3D model in the real environment. During undergraduate study, a component-level interaction system in an augmented reality (AR) had been developed [4]. The system presented real-time interaction in a marker-based AR system; 3D components virtual object is registered on the marker and manipulates each component of the virtual by using traditional input devices keyboard (see Figure 1.1).

In this system, we are aware that input hardware such as keyboard and mouse do not allow 3D interaction in an AR application in a natural manner. These input methods are limited in depth distance movement and inhibited in the degree of freedom. Besides, the presence of markers is invasive and unnatural in a real environment.

It is possible to achieve a more natural interaction with a virtual object by expanding the dimension of the interaction. Thus, in this research, we proposed the utilization of two handed gestures, where one of the hands is outstretched for registration of 3D virtual objects in the real environment while the other hand serves as an input to interact directly with the virtual objects.

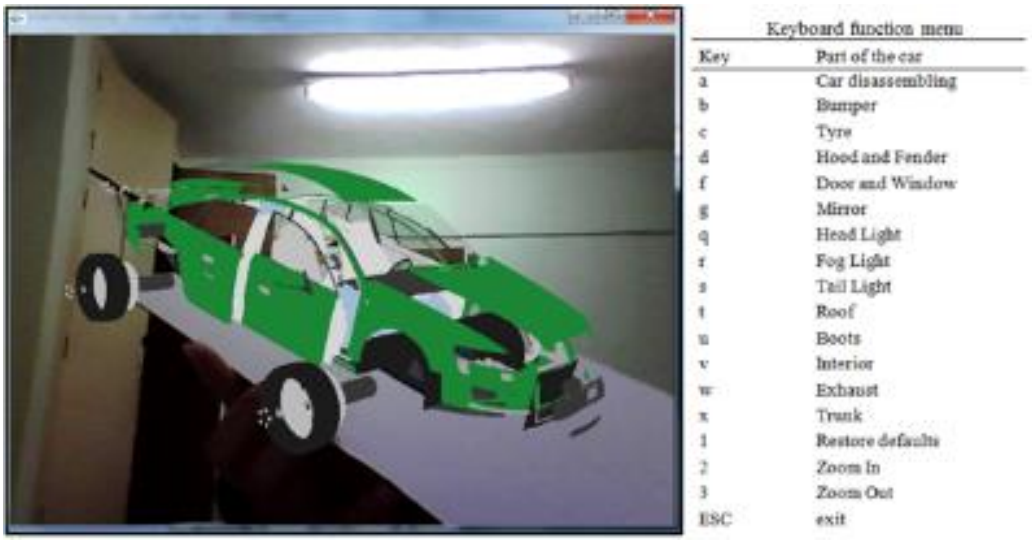


Figure 1.1: All components of a virtual object are displayed through a marker and controlled by a keyboard

1.3 Research Objectives

This study embarks on the following objectives:

- To enhance watershed transform algorithm for solving over-segmentation problem.
- To develop two homogeneous object tracking system under mutual occlusion.
- To purpose a cost-effective method for two hand tracking system under mutual occlusion.

1.4 Problem Statement

The key attraction in Augmented Reality (AR) is its fundamental capability to enable the virtual and real objects to co-exist in the real environment. However, human's adventurous nature does not satisfied with only the co-existence status. The urge to interact with the virtual objects in the real environment has caused further research work and resulted in interactive AR. Typically, interactive AR involves the interaction between two entities, input device and the displayed virtual object. To achieve a natural interaction in AR environment, two-handed interaction are highly preferred, that is by using an outstretched hand for registration of 3D virtual objects,

while another hand as the input device for interaction. However, when two objects that are spatially separated in 3D may occlude each other in the projected 2D image plane, and causing two-handed gestures unable to be recognized directly by vision based. In this thesis, we present a solution for two-hand occlusion by using watershed transform. The main idea is to start from a two-hand occlusion image in binary format, then form a grey-scale image based on the distance of each non-object pixel to object pixel. The watershed algorithm is applied to the negation of the grey scaled image to form watershed lines which separates the two hands.

1.5 Organization of Thesis

The main body of this thesis is divided into six chapters and each chapter discusses different issues pertaining to the research. Below are the outlines for each chapter:

(i) Chapter 1 - Introduction

This chapter provides the background information about HCI which includes the motivation of this research, the problem statement, and research objectives.

(ii) Chapter 2 - Literature Review

Some related literature and researches are reviewed and discussed in this chapter.

(iii) Chapter 3 - Watershed Transform

In this chapter, watershed transform based on morphological approach will be presented which includes a brief review on the basic definitions and various morphological tools. A two-handed segmentation using watershed transform is also demonstrated here.

(iv) Chapter 4 - Methodology

Analytical methods and the procedure of this project development are discussed.

(v) Chapter 5 - Result and Discussion

This chapter is a follow-up to the results from the previous chapter. It presents the output results from the proposed watershed transform approach and they are compared with the morphological method.

(vi) Chapter 6 - Conclusion

A complete summary of the research and future work are set out in this chapter.

CHAPTER 2

LITERATURE REVIEW

2.1 Hand Gesture Recognition

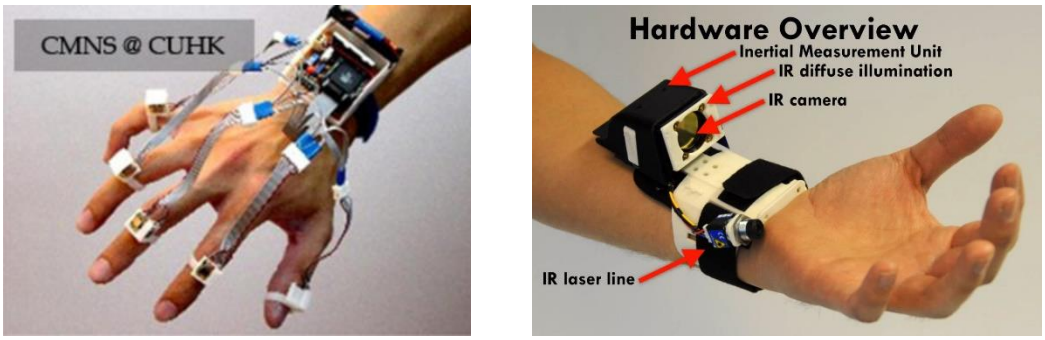
Hand-interaction mechanism is popular in many HCI applications and has been widely studied since long ago. The ultimate aim of the human computer interaction is to deliver a more natural interaction between human and computer which can be achieved by incorporating gestures in HCI is currently a trending research area. Gesture recognition refers to the entire process of tracking human gestures to their representations and converting them to semantically meaningful commands [5]. The goal of research in hand gesture recognition is to develop a system to identify the human gestures explicitly as input and process these gesture representations for device control through mapping of commands as output. Some of the ways to detect these gesture information are sensor-based, vision-based, and hybrid tracking techniques. The reviews of sensor-based and vision-based techniques are presented in this chapter.

2.1.1 Sensor-Based

In sensor-based technique, a device is employed to obtain the data about the position and orientation of hand gestures and its movements. There are many different types of sensor that have been developed. For example, Figure 2.1(a) show the magnetic-sensing devices for finger joints that transmit complete signals for hand gesture recognition [6], Figure 2.1(b) infrared (IR) sensors that are mounted on the fingers tips and joints to detect their orientations and positions to recognize the hand gestures [7]. The accuracies of gesture detection by these techniques are high.

However, the devices involved are expensive and their attachments to the fingers limit hand movements, thus not allowing natural interaction.

Kinect sensor has emerged a new hand gesture recognition method in the recent development of depth cameras in a lower cost. Rather than wearing a data glove, Kinect motion sensors are able to track the hand movements more accurately. Kinect sensors are motion-sensing input device based on skeletal tracking and widely used in articulated head recognition [8], human body tracking [9], human action recognition, and hand movement tracking [10]. However, there is no specific tracking for the hand at finger level for gesture recognition; only the position of the hand, not the fingers, is detected in Kinect. Thus, a little work needed for Kinect to detect the details at the level of individual fingers.



(a) (b)

Figure 2.1: Sensor based hand tracking; (a) Magnetic-sensing devices; (b) Infrared sensor

2.1.2 Vision-Based Hand Tracking

In order to keep the hand tracking system in low cost and to circumvent the problem of hand movement limitations in sensor-based systems, vision-based hand tracking is proposed in our research. Vision-based hand tracking involves recording of the hand movements using video cameras, where the gestures in the images are recognized and then analysed by image processing algorithms [11]. Hand gesture recognition is known to be challenging due to the environmental properties and system

requirements involved. The main challenges of the vision-based hand gesture recognition development are depicted as follows [12]:

a. High-dimensional problem: Studies have shown that a hand motion involves at least six dimensions. Even only taking into account that a natural hand motion involves less than 20 degrees of freedom (DoF), there are still many parameters need to be estimated due to the independence between fingers.
b. Self-occlusions: It is very difficult to compute the hand gesture perfectly because there are a huge number of possible hand shapes could be estimated with many self-occlusions depending on the viewpoints.
c. Processing speed: During run time, great amount of data has to be processed for each input image. An efficient algorithm and high performance hardware are recommended so that the system able to run in real time.
d. Uncontrolled environment: The extraction of a static object from a simple background is not a straight forward task. It is even more so when segmenting a dynamic hand from the real word environment with different lighting or background.
e. Rapid hand motion: The supported frame rate and the tracking algorithms employed in the most common commercial cameras are not powerful enough to capture high-speed hand motions during run time.

It is nearly impossible to develop a system that fits for all purposes. Furthermore, most of the latest developed systems do not overcome all the above mentioned challenges and certain kind of restrictions have been applied on the input image. Some common restrictions on the environments are the assumptions that the background is

static or simple and that the hand is the only skin-coloured object. The hand gestures that are supported by the system may also be limited to those which are simple and of low degree of freedom. Vision-based hand gesture recognition is very system dependence application, different system may have different impact from different restrictions due to the nature and implementation of it. Generally, the restrictions which cause less impact to the system would be implemented to improve the accuracy of the hand gesture recognition.

Currently, the two main branches of the vision-based hand gestures recognition are 3D model-based approaches and appearance-based approaches [13]. The former approach relies on the 3D kinematic hand model with considerable degree of freedom, and the hand parameters can be acquired by comparing the input images with the possible 2D appearance calculated by the 3D hand model. This approach ideally provides realistic interactions in virtual environments. However, this approach is difficult to carry out in real time because it typically uses overcomplicated and heavy algorithms to extract the exact joint angles of hands.

On the other hand, appearance-based algorithm is known to be good in dealing with object recognition. A statistical model of various object appearances or so-called template is used for the object recognition task and it is stored in a database. The hand gestures are simulated by matching the appearance of simulated hand gesture to the appearance of the set of predefined template gestures [14]. Appearance-based approaches perform well in real time simulation as their 2D image features are easily utilized. Lately, much effort has been paid on appearance-based methods. Among all the appearance-based methods, the simplest and frequently used approach is by looking for skin-coloured regions in the image [15]. Nevertheless, this comes with some trade-offs. First, skin-colour detection is very sensitive to lighting conditions. In

spite of practical and efficient methods exist for skin-colour detection under controlled (and known) illumination, the issues of building a flexible skin model and implementing it over the simulation time remains challenging. Second, it only works for the inputs without any other skin-like object in the background. A study had been done by Lars and Lindberg [16] using scale-space colour features to recognize hand gestures. The fundamental of this gesture recognition technique is based on feature detection and it is user independent, but the study presented only real-time simulation without any other skin-coloured object appear in the background. Although skin-colour detection is practical and efficient under strictly-controlled working environments, but it cannot distinguish two hands apart when they are overlapped on each other in the image.

A few researches had been carried out to address the vision-based occlusion problem for two hands that are spatially separated. For example, modelling hand gestures from the Hidden Markov Models was applied to resolve the occlusion problem by fixing multiple cameras and select the 'best-view' image by filtering out other occluded images [17]. In [18], the crossing-hand occlusion problem was resolved by tracking the arms' motions with template matching and rotating the template to find changes in occlusion dissimilarities, which allowed differentiation of the hands. In [19], the approach in the two-hand occlusion problem was by modelling the spatial synchronization of bimanual movement using velocity and acceleration of each hand. Although this method was claimed to be able to handle all possible cases of occlusion, it only tracked the region of interest and the hand boundary was unclear. [20] Made blob-like patterns from the image and hand gestures were detected using the blobs. It was efficient in two-hand tracking, but only limited to hand gestures.

2.2 Image Segmentation Method

Image segmentation holds an important role in computer vision and image analysis. Practically, researchers are only interested in particular components of an image. These components are often referred as a target objects generally correspond to the image in a specific and unique nature of area. It has to be extracted and partitioned for identification and analysis. By dividing an image into regions based on different features such as greyscale pixel, colour, and texture and so on, this process is known as image segmentation.

A lot of segmentation methods had been studied previously, and categorization is necessary to outline the methods properly. Most of the earlier literature had categorized the segmentation methods as below [23]:

a. Pixel-based segmentation is the method of dividing a digital image into multiple segments or groups of pixels. Pixel-based skin colour detection segmented method is the most common method used in vision based hand recognition, which classify each pixel as skin or non-skin individually. Several colour spaces have been adopted in the previous studies, which include RGB, normalized RGB, YCbCr, HSV,etc.

   RGB colour space was found from CRT display application. It is commonly used for handling and storing digital image data because of its simplicity to describe colour in a combination of 3 coloured rays which are red, green and blue. Nonetheless, high correlation among channels, significant perceptual non-uniformity, mixing of chrominance and luminance data leads RGB to be a non-favourable method for colour analysis and colour based recognition algorithms. In order to overcome this problem, several colour spaces were introduced which allow user to specify the colour properties numerically, for

instance, HSV, Hue Saturation and Intensity based colour spaces. These colour spaces describe colour with intuitive values, based on the artist's idea of tint, saturation and tone. Hue uses colour such as red, green, purple and yellow to define the dominant colour of an area, saturation calculates the ratio of the colourfulness of an area to its brightness. Whereas, the intensity based is related to the colour luminance. These colour spaces are accepted widely especially in skin colour segmentation field due to its intuitiveness of colour space components as well as the explicit discrimination between luminance and chrominance characteristics. It is notices that, these colour spaces have disadvantages including hue discontinuities and the computation of brightness, which conflicts badly with the properties of the colour vision and the polar coordinate system of Hue-Saturation spaces, yields a cyclic nature of the colour space which is not suitable for parametric skin colour models that requires tight cluster of skin colours.

b. Threshold-based segmentation. This segmentation method is a combination of histogram thresholding and slicing techniques. This method may be applied on an image directly; either combined with pre-processing or post-processing techniques.
c. Edge-based segmentation. Edges detected in an image are assumed to represent object boundaries which are used for object recognition.
d. Region-based segmentation. This segmentation method examines neighbouring pixels of initial seed points and decides whether these pixels should be added to the region. The regions are then spread from these seed points to adjacent points based on a region membership criterion. The criterion could be pixel intensity, greyscale texture, or colour.

However, all the basic methods mentioned above are not able to segment overlapping object or images under occlusion. Therefore, many researches were carried out to find a solution to this problem. Most of the common methods, such as in [24], involve identification and labelling of occlusion boundaries based on prior shape segmentations. In [25], the author developed Bayesian segmentation algorithm that uses a region-based background subtraction. In [26], Kalman filter-based applied active contour model in non-rigid object tracking that can deal with the occlusion problem.

Most of the methods above would fail if the moving objects change their appearances substantially and thus not applicable in two-hand tracking.

s

CHAPTER 3

WATERSHED TRANSFORM

This chapter briefly explains the way watershed transform deals with image segmentation problems by means of mathematical morphology. Watershed transform is based on grey-scale mathematical morphology. It has been extensively used in various kinds of image segmentation especially in getting the segments in low-contrast and weak-boundary regions. Watershed transform is a better option because it always produces closed contours, which is very useful in image segmentation.

The review of watershed transformation, its basic notions and various morphological tools will be presented in the next sections. Furthermore, a two-handed segmentation by watershed transform is demonstrated in the following sections.

3.1 Principles of Watershed Transform

The fundamental of watershed segmentation is interpreted as a topographic surface, where the value of each pixel represents the elevation of the particular point. High surface elevation is represented by the greater grey intensity value, while low surface elevation is represented by the lesser grey intensity value (see Figure 3.1(a)). In topographic representation, images are illustrated in three basic components, namely the minima, catchment basins, and watershed lines. Each of the local minimum value is assumed to be a catchment basin and the basin's boundary as watershed lines. Figure 3.1 sets out the graphical explanation of these terms.

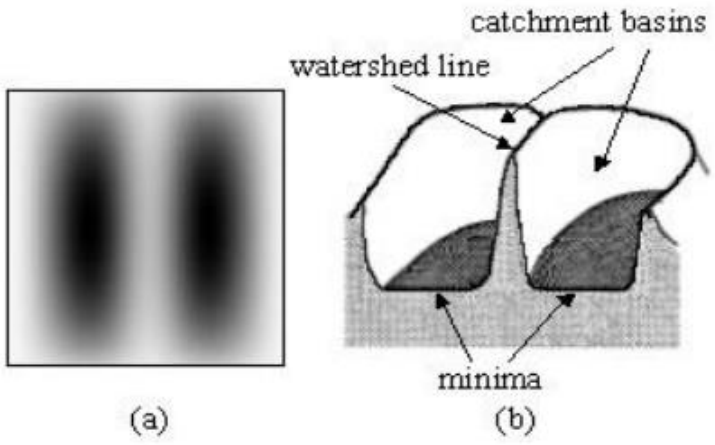


Figure 3.1: Graphical explanation of watershed transform; (a) grey-level image. (b) Topographic surface of (a)

Eventually, the key intention of watershed transform is to acquire the watershed lines which have the highest grey intensity value in an image. From Pierre Soille and Luc Vincent's explanations, watershed segmentation of the grey-scaled image is performed by using the immersion process [27]. This process can be visualized and explained as a concave topographic surface at each local minimum filled by the water, ultimately the surface will be gradually immersed in water. Immersion is started from the minima of the lightest grey intensity point in an image, the water will fill up different catchment basins of topographic surface to form 'lakes' at the same rate. Then, a 'dam wall' will be established to avoid all the 'lakes' on the topographic surface from merging and overlapping each other. When the topographic surface is flooded to its highest level, the 'dam walls' will eventually form the outlines of the watershed lines. Figure 3.2 depicts this procedure graphically:

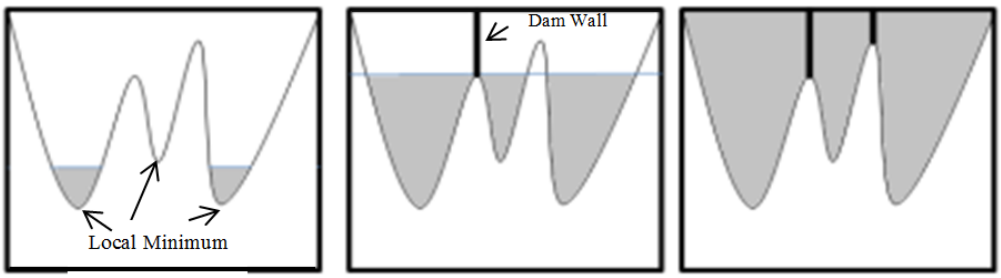


Figure 3.2: Three different stages of watershed construction by flooding process

3.2 Mathematical Expression of Watershed Transform

Let define a two-dimensional grey-scale image as,

- I where its definition domain is denoted as $D \subset Z^2$
- h is the grey-value of image, I.
- h_{min} and h_{max} are the smallest and the largest values taken by I in domain D, respectively.

In the following, $T_h(I)$ expresses the threshold of I at level h [28]:

$$T_h(I) = \{(p)\epsilon D_I | I(p) < h\}.$$

Assume M_i, where $i = 1, \dots, h$, be the set of points in the regional minima (catchment basins) of image I, and $C(M_i)$ be the point of catchment basins associated with the regional minimaM_i. Let $C_h(M_1)$ be the set of points in the catchment basin associated with M_1 that are flooded at stage h [29]:

$$C_h(M1) = \cap \{C(M1), T_h(I)\}.$$

The watershed of I is the set of points which do not belong to any catchment basin [29]:

$$watershed = D \cap (\bigcup C(M_i))^c$$

3.3 Basic Notions of Mathematical Morphology

Mathematical morphology mostly deals with the mathematical theory in describing shapes using sets. All mathematical morphology operations are based on dilation and erosion. Both dilation and erosion operations are produced by the interaction of a set structuring element with a set of pixels of interest in the image. The structuring element has a shape and an origin that are used to determine the precise details of the effect of the operator on the image.

Let a grey image be represented by a scalar function$I(x, y)$. The dilation δ and erosion ε of the image by the structuring element $A(x', y')$ is defined by [28]:

$$\delta_A = (I \oplus A)(x, y) \coloneqq \sup\{I(x - x', y - y') | (x', y') \in A\},$$

$$\varepsilon_A = (I \ominus A)(x, y) \coloneqq \inf\{I(x - x', y - y') | (x', y') \in A\}.$$

Dilation can be expressed as the expansion of image by the shape whilst erosion as the contraction of the image size. Using these notions, Beucher identified where the edge can be detected by examining the difference between dilation and erosion of an image by the elementary structuring element, called morphology gradient *g(I)* [28].

$$g(I) = \delta_A(I) - \varepsilon_A(I)$$

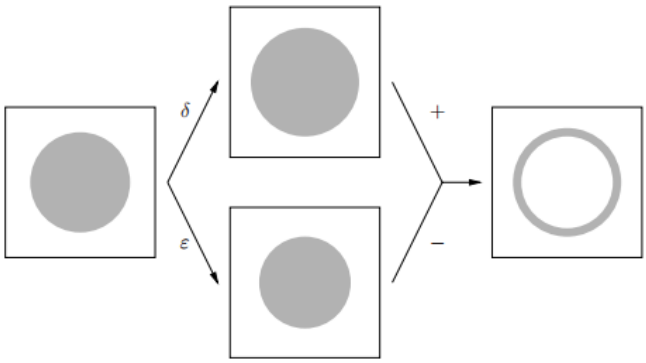


Figure 3.3: Edge detection by taking the difference of the dilation and erosion of an image

The gradient sticks out on both sides of the actual edges which decomposed into two-half gradients where the inner gradient $g^-(I)$ adheres to the inside of object, and the outer gradient $g^+(I)$ adheres to the outside of object[28]:

$$g^-(I) = I - (I \ominus A)$$

$$g^+(I) = (I \oplus A) - I$$

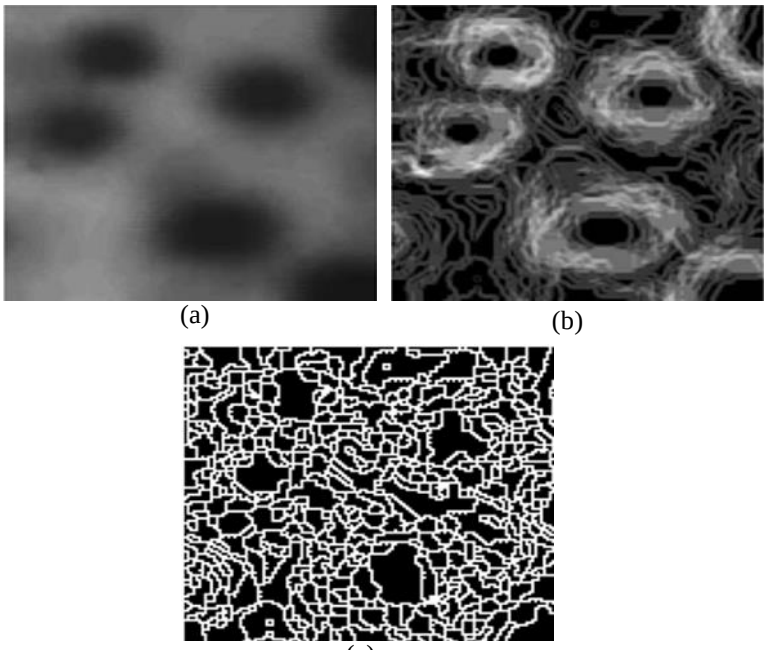
$$g(I) = g^+(I) + g^-(I)$$

The idea is to apply a dilation process close to a maximum and an erosion process in the vicinity of a minimum in order to identify the region whether it is belonging to the influence zone of the maximum or the minimum. Kramer and Bruckner have proposed the morphological Laplacian, ∇I [28]. A morphological Laplacian ∇I, is given by

$$\nabla \boldsymbol{I} = (\boldsymbol{I} \oplus \boldsymbol{A}) - 2\boldsymbol{I} + (\boldsymbol{I} \ominus \boldsymbol{A}).$$

If $\nabla \boldsymbol{I} < 0$, the region is considered as the influence zone of a maximum, while if $\nabla \boldsymbol{I} > \boldsymbol{0}$, the region is the influence zone of the. $\nabla \boldsymbol{I} = \boldsymbol{0}$ Is interpreted as the edge location.

As the result, the region of segmentation can be identified by distinguishing the influence zones of minima and maxima. Figure 3.4(c) shows a watershed line obtained from an image gradient. Each connected region contains one local minimum in the corresponding gradient image.



(a) (b)

(c)

Figure 3.4: Watershed line obtained by image gradients; (a) input image; (b) image gradient; (c) watershed line of input image

3.4 Literature Review of Watershed Transform

The original algorithm of watershed transform is vulnerable to noise, which may yield miscellaneous over-segmentation where a large number of segmented regions are formed. To avoid over-segmentation, many improved approaches have been proposed. Most of the approaches can be classified as pre-processing and post-processing. Since watershed algorithms are extremely vulnerable to noise, image pre-processing such as median filter, colour morphology and distance transform can be further applied to remove noise. However, post-processing merges some of the basins in an appropriate way by eliminating irrelevant watershed lines.

Various approaches to improve the watershed transform may be found in literatures, such as pre-processing of image such as distance transform and marker based watershed transform. For example, in [30], the author presented a watershed algorithm incorporating morphological opening-closing filtering and distance transformation to segment image in his study. This approach effectively addressed the over-segmentation issues in segmenting dowel images, and output a better and satisfactory segmentation result compare to the use of stand-alone distance transformation algorithm. In this proposed algorithm, the threshold of an image pixel is predicted by calculating the mean of the greyscale values of its neighbouring pixels. The square variance of the greyscale values of the neighbour pixels are also calculated as an additional determining condition, so that the output of the proposed algorithm is the edge of the image. Indeed, the proposed algorithm is equivalent to an edge detector in image processing. In [31], the author presented watershed transform as a strong morphological tool for image segmentation. Yet it causes over-segmentation and noise in the image produced. So, they used marker-based watershed technique to reduce noise and over-segmentation. Marker watershed segmentation method first performs

bilateral filtering for image processing to eliminate the noise effect in the post-processing, followed by the use of distance transform and shape reconstruction method for image processing, resulting in a more precise positioning profile. In [32], the author focused on segmentation of touching cell images which is helpful in recognizing morphological structural model of touching cells. A morphological reconstruction approach is used markers based on watershed transform. The Internal and external markers are used to mark the regions of the image. Then, watershed segmentation algorithm was applied and once the results were obtained, it was compared to that of traditional techniques. The new approach was found to be better than the old one.

Over-segmentation problem remains in most of the watershed segmented images, even though they have been pre-processed. To overcome this problem, numerous region-merging methods have been proposed for watershed post-processing. For instance, [33] proposed wavelet-based watershed image segmentation technique that managed to solve the over-segmentation issue at the same time provide noise suppression, unfortunately it failed to implement because the low-contrast edges exist within the regions of interest. Besides that, [34] proposed multi-scale gradient watershed image analysis, and [35] had formulated partial differential equations for image de-noising or edge enhancement.

3.4.1 Method Based on Distance Transform

The distance transform base watershed approach is a generally used and improved algorithm to separate touching objects. There are only two grey levels 0 and 1 present in a binary image and these values represent black and white colour respectively. In the case of two blobs were merged together in a binary image, only one minimum and a catchment basin will be formed in the topographic surface. In order to segment the connected blobs using watershed transform, distance transform

has to be implemented first to convert a binary image to a grey scale image with the grey tones running from the boundary to the centre of the object based on the distance values. This grey scale image is then ready for watershed transform. Figure 3.5 sets out the result of distance transform by using Figure 3.5(a) as input image. Figure 3.5(b) shows the distance of each pixel, running from white to black, where white being the pixels with 0 distance to the object and black being the farthest from the object. Figure 3.5(c) presents the negation of Figure 3.5(b).

Watershed transform is described as a topographic representation of negation of distance transform map (see Figure 3.5(c)). The greater the pixel values, the higher the gradient of the ridge area in image while low-gradient pixels are considered as basin. Watershed transform could be illustrated as a water stream falling from the ridge. The water flows along a path to reach the catchment basin to form a region. The watershed lines divide the catchment sink into different areas to form the object contours in the image.

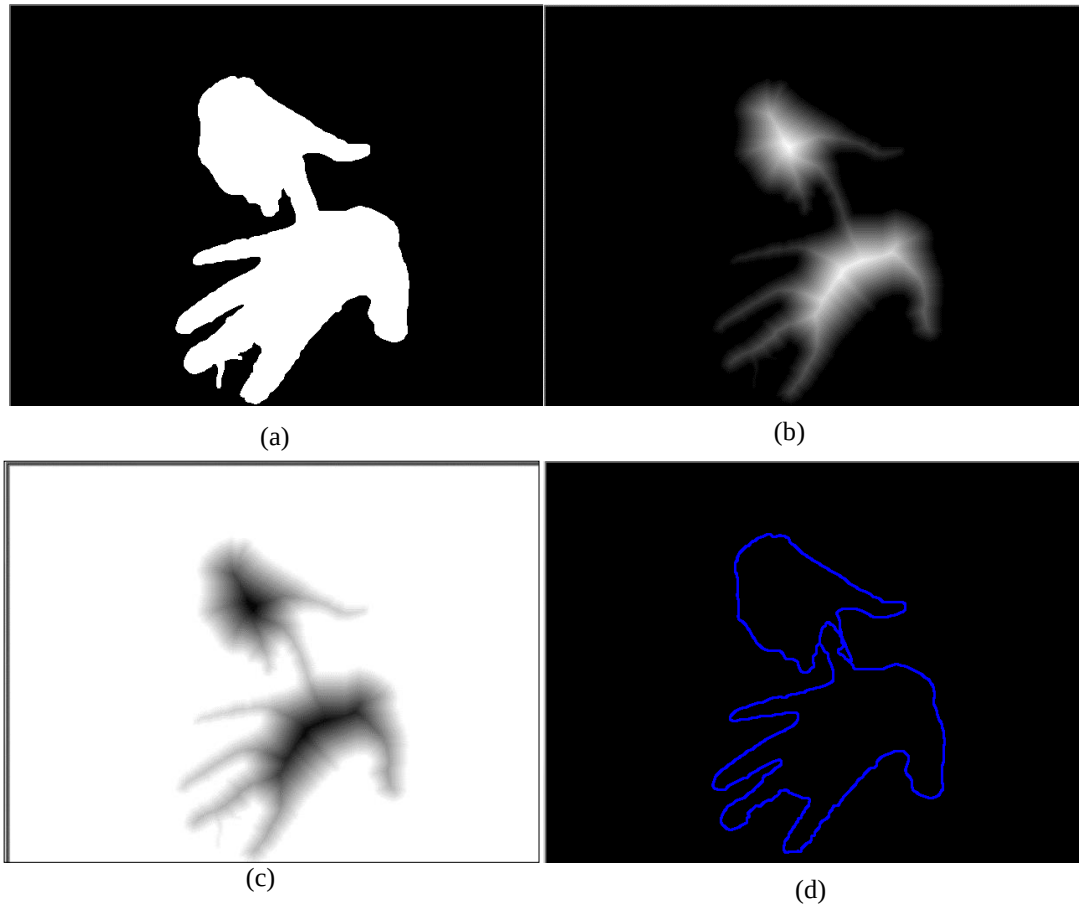


Figure 3.5: Distance transform for watershed; (a) input image; (b) result of distance transform applied to input image; (c) negation of (b) result of watershed transform based on distance transform

3.4.2 Marker-Controlled Methods in Watershed Transform

Marker-based watershed transform was introduced by Meyer [36] to effectively segment an image. The marker is used to limit the number of regions by specifying the objects of interest and generate starting points for the flooding process. These starting points are considered as centre of points of the catchments. Since there is exactly a region created per marker during the flooding process, the use of lesser markers is encouraged to lower down the number of regions in the final segmentation. A common choice for the marker is on the peak of object; it can be a point, a line, or a region. The marker can be extracted manually or by another segmentation algorithm. As discussed in the previous section, the greater the pixel values, the higher the gradient of the ridge area in the image. Distance transform could be used to locate the peaks of the object and mark them as regions of interest.