AUTOMATIC LANDMARKING
ON 2.5D FACE RANGE IMAGES

Pui Suk Ting

Master in Advanced Information Technology
(MAIT)
Declaration

I hereby declare that this master’s thesis is my original work. I have not copied from any other students’ work or from other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person. All content and ideas drawn directly or indirectly from external sources are indicated as such. The thesis has not been submitted to any other examining body and has not been published.

Signature:

Name: Pui Suk Ting

Date: 6 June 2014

Matric Number: 11031854
Acknowledgments

Firstly, I would like to express my sincere appreciation to my supervisor, Dr. Jacey-Lynn Minoi, for her guidance and patience throughout this project. Her constantly guide along the way has helped me completed this project successfully.

I offer my regards and blessings to all of those who offered a generous help in any respect during the completion of this work. Last but not least, my warmest thanks to my family who are always my strongest supporters. Their kindness and love were the main strength for me over the years. Thank you for believing in me.

This research was awarded and funded by ScienceFund under the Ministry of Science, Technology and Innovation (MOSTI).
Abstract

This project presents a novel approach for automatic landmarking process on face range images. The approach consists of feature extraction and feature localisation methods. In recent years, methods to improve existing face processing applications have increased rapidly. This includes automatic landmarking method which could be an important intermediary step for many face processing applications, such as for face recognition, face analysis and etc. The approach aims to locate facial feature points automatically, such as the nose tip, the mouth corners, chin, etc., without the intervention of human. Automatic landmarking holds a number of added advantages over manual landmarking especially if dataset is large, the landmark selection would be less time consuming. Identifying features on a face automatically may be a challenging process for computing. Our human vision system can perceive salient feature easily without any difficulties. For instance, a human is able to detect the eyes, the nose tip and/or the mouth of a person at a first glance. However, a computer system is unable to do such task easily and effortlessly. Therefore, a method to automatically detect and label landmarks on the features of the face is developed. Firstly, features or primitive surface types are extracted from range images. The primitive surfaces are derived using Mean (H) and Gaussian (K) curvatures from a down-sampled by Gaussian Pyramid approach. Otsu’s algorithm is used to place landmark on the extracted facial features and/or regions. In summary, we have successfully implemented an automatic landmarking method and an interactive tool has been developed to ease the visualisation of the overall processes.
Abstrak

**Paper Contributions**


**Pui, S. T., & Minoi, J. L.**: Towards automatic landmarking on 2.5D face range images. In Information Technology in Asia (CITA), 8th International Conference, pages 141-146, IEEE, July 2013, Kuching, Malaysia.
# Table of Contents

**Chapter 1: Introduction**  
1.1 Introduction  
1.2 Previous Works on Automatic Face Landmarking  
1.3 Motivations  
1.4 Challenges  
1.5 Aims and Objectives  
1.6 Overview of Report  

**Chapter 2: Face Landmarking Techniques**  
2.1 Introduction  
2.2 Quality Improvement through Pre-processing  
  2.2.1 Illumination Artifacts Removal  
  2.2.2 Noise Removal Techniques  
  2.2.3 Holes Removal Techniques  
  2.2.4 Geometric Corrections Techniques  
2.3 Face Landmarking Techniques  
  2.3.1 Model-based Methods  
    2.3.1.1 Explicit Methods  
    2.3.1.2 Implicit Methods  
  2.3.2 Texture-based Methods  
    2.3.2.1 Transform-based Methods  
    2.3.2.2 Template-based Methods  
2.4 Face Landmarking on 3D Images  
  2.4.1 Mean (H) and Gaussian (K) Curvatures
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Summary of quality improvement and face landmarking methods.</td>
<td>23</td>
</tr>
<tr>
<td>3.1</td>
<td>The detail of H&amp;K curvatures primitive surfaces and the colour labels.</td>
<td>34</td>
</tr>
<tr>
<td>3.2</td>
<td>The size of the range image at each pyramid level</td>
<td>36</td>
</tr>
<tr>
<td>4.1</td>
<td>Primitive surfaces and the colour labels</td>
<td>47</td>
</tr>
<tr>
<td>4.2</td>
<td>The summary on the number of landmark labels on the primitive surfaces during the landmarking process.</td>
<td>53</td>
</tr>
<tr>
<td>4.3</td>
<td>Accuracy results by comparing automatic and manual landmarking points.</td>
<td>61</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Categorization of landmarking approaches (Çeliktutan et al., 2013). 13
3.1 An overview of the automatic landmarking framework 26
3.2 The convolution of scale space with Gaussians to produce Difference of Gaussian (DoG) images in each octave (Lowe, 2004). 28
3.3 Keypoint detection in scale-space. 29
3.4 The sign of p based on differential geometry on a torus. 32
3.5 Dome; K > 0 32
3.6 Cylinder; K=0 33
3.7 Saddle; K<0 33
3.8 Decomposition of the Mean and Gaussian curvatures over facial model (Szeptycki, 2012). 34
3.9 Example of illustrating a Gaussian pyramid (Lo, 2009). 35
3.10 An overview of the feature localisation framework 36
3.11 Centroid, C of common shape. 39
4.1 The interface of the developed interactive extraction tool 42
4.2 Gaussian smoothing is not applied (a1) and applied (b1) respectively, enlarged face area (a2 and b2) on range image set 1. 44
4.3 Gaussian smoothing is not applied (a1) and applied (b1) respectively, enlarged face area (a2 and b2) on range image set 2. 45
4.4 Gaussian smoothing is not applied (a1) and applied (b1) respectively, enlarged face area (a2 and b2) on range image set 3. 46
4.5 The extracted primitive surfaces (in a1, b1 and c1) and the enlarged image on the highlighted areas (in a2, b2 and c2), respectively, from three different
range images.

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6</td>
<td>Extracted facial features (a) and binary conversion (b) on range image set 1.</td>
<td>49</td>
</tr>
<tr>
<td>4.7</td>
<td>Extracted facial features (a) and binary conversion (b) on range image set 2.</td>
<td>50</td>
</tr>
<tr>
<td>4.8</td>
<td>Extracted facial features (a) and binary conversion (b) on range image set 3.</td>
<td>51</td>
</tr>
<tr>
<td>4.9</td>
<td>Landmark labels on the extracted primitive surfaces on range image set 1.</td>
<td>55</td>
</tr>
<tr>
<td>4.10</td>
<td>Landmark labels on the extracted primitive surfaces on range image set 2.</td>
<td>56</td>
</tr>
<tr>
<td>4.11</td>
<td>Landmark labels on the extracted primitive surfaces on range image set 3.</td>
<td>58</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Introduction

Over the last decades, 3D images have become popular due to the advancement of 3D sensor and camera technologies; alongside with 2.5D range images. Range images have a number of added advantages over 2D images. A 2.5D image is defined as a simplified three-dimensional (x, y, z) surface representation that contains at most one depth (z) value for every point in the (x, y) plane (Vaughan, 2011). One can think of a 2.5D image as a grayscale image, whereby a black pixel corresponds to the background, while the white pixel represents the surface point that is nearest to the camera (Abate et al., 2007). The 2.5D face images enable depth perception and allow one to manipulate the image like the 3D image. In addition to range data, colour perception on a face image is also possible. The various sources of information from the range image can be extracted to help derive different features regions. 2.5D range images are thus used in this research as a dataset to define the keypoint descriptors by extracting the facial surface information.

In this project, we aim to develop an automatic landmarking method on face data set using 2.5D range images. Automatic facial landmarking is a vital process, which could be employed on any face applications, for analysis, registration and recognition. Landmark is defined as the voxel in the image, which corresponds to a specific point on the anatomy of the object (Talairach & Tournoux, 1988). Landmarking process of a face is described as the detection and localisation of distinct face feature points (Çeliktutan et al., 2013). Automatic landmarking allow facial features to be located and labelled without the intervention of human (Tie & Guan, 2013).

Unlike the conventional manual landmarking, whereby a human intervenes to locate facial features visually, automatic landmarking holds a number of advantages and is able to
locate features better and may be accurately. This approach could save time especially when the face data sets are getting bigger and more facial features need to be landmarked.

Automatic landmarking method is usually challenging and computationally complex. Until today, there is no available automatic landmarking method for face data sets. Our human vision system and brain mechanisms can quickly perceive distinct features, such as the edges, the tips or corners of any object without difficulties, whereas a computer is unable to do such tasks easily and effortlessly (Çeliktutan et al., 2013). A face image contains complex features with a large degree of background and face variations, such as the identity of the person, facial expressions, head poses, illumination, facial hair variability, age, gender, cluttered background and so on. In addition, these images are stored in different format types, resolution, sizes, scales and rotations. These variants lead to problems and issues to developing an automatic landmarking system.

1.2 Previous Works on Automatic Facial Landmarking

In recent years, automatic landmarking process is becoming an urgent interest with the increased of the data sets and the development of face applications. It benefits applications ranging from face registration to face and facial expression recognition (Szeptycki et al., 2009). Identifying landmarks for registration and pre-processing is necessary in any face applications, and it is usually done manually. Therefore, in making any face systems fully automatic from acquiring faces to processing those faces, automating landmark selection process in the registration method is imperative (Perakis et al., 2009a).

A number of researchers have experimented different techniques and methods to detect and select landmarks on face images. Most of these existing works use prior knowledge computed from facial geometry-based analysis and curvatures analysis (Szeptycki et al., 2009).

The most common and popular method for extracting facial features is based on the
curvatures approach (Szeptycki, 2012), which has been widely used since the 1980s. Besl (1986) introduced Mean (H) and Gaussian (K) curvatures method to segment facial surface into eight different surface types. Other researchers (Trucco & Fisher, 1995; Gordon, 1992; Tanaka et al., 1998; Kim et al., 2001; Moreno et al., 2003) also adopted this approach on range images.

Shape Index (SI) method introduced by Koenderick and van Doorn (1992) decouples the shape and the magnitude of the curvedness of the features. In a series of publications reported in Colbry et al. (2005), Lu & Jain (2005), Lu & Jain (2006), Lu et al. (2006), and Colbry (2006), these researchers presented methods to locate the positions of the corners of the eyes and mouth, the nose and also the tip of the chin. They developed heuristic methods to identify the tip of the nose more efficiently. The candidate landmarks were filtered by using a statistical model of landmark positions.

Dibeklioglu et al. (2008) developed an automatic 3D facial landmarking algorithm by using statistical and heuristic approaches. The employed statistical landmark localisation was based on the analysis of the detected local features. Heuristic method using the curvature method was used to localise the tip of the nose under several variant conditions. This algorithm can be employed to model any facial landmarks, but face data with pose variation in testing and training sets were not taken into consideration.

Nair and Cavallaro (2009) presented a method to detect candidate facial landmarks on 2.5D image scans. They applied the Shape Index (SI) and the curvedness index to extract feature points on the nose tip and the corners of the eyes. The feature points are fitted to the remaining data set using the detected three control points/landmarks for the registration process. However, the method cannot be applied to missing data in pose self-occlusion.
In Perakis et al. (2009a) and Perakis et al. (2009b), the authors presented methods for better detecting facial landmarks on 2.5D image scans. The selected candidate landmarks are the inner and outer corners of the eyes, corners of the mouth, the tip of the nose and the chin. In order to locate these candidate landmarks, local shape and curvature analysis were applied using SI method, extrusion maps and spin images. The landmarks are identified by matching the located landmarks with the facial model landmarks using a statistical model. Curvature index model is commonly used in feature extraction; hence our project will employ this model.

1.3 Motivations

As prior mentioned, a fully automated face recognition applications would require the landmarking of points to be fully automated (Perakis et al., 2009a). The commonly used feature extraction method is the curvature-based approach. However, this approach still contains a problem on images with missing data, variant to scale, rotation and illumination, resulting to failures in extracting the required features and locating landmark points precisely.

Therefore, it is our aim to improve the feature extraction approach invariant of scale and orientation, for accurate landmarking of points.

1.4 Challenges

Illumination, pose, occlusion and facial expression are the common face variations in images that could affect the accuracy of the feature extraction and landmarking methods (Creusot, 2011). Images usually suffer from irregular or poor illumination. To handle the illumination problem, 3D face information can be used. In addition, 2.5D images are also illumination invariant whereby the extra dimension information or shape information is independent from any illumination conditions (Liu et al., 2009). In comparison to 3D data, 2.5D holds better advantages as it also contains depth information that gives 3D information and the data set size is considerably smaller, which could speed-up any complex processing.
Pose variation is another challenge in images processing. For instance, in locating the exact tip of the nose on a profile face would be easier than on a frontal face. Unlike 3D face data that is invariant to pose, in order to cope with pose variation in 2D and 2.5D images, these images need to be pre-processed and normalized (Conde & Serrano, 2005).

Curvature-based approach may be a promising method as a pre-processing method to eliminate surface irregularities (Dibeklioglu et al., 2008) or as a feature extraction method but this approach is still variant to scale and orientation.

1.5 Aim and Objectives

The main aim of this research is to develop a method to landmark points on the facial features from the 2.5D face range images automatically.

The objectives include:

1. To investigate the general threshold values from the developed feature extraction method to determine the candidate facial landmarks such as the tip of the nose, the eye and mouth corners, etc.

2. To develop a feature localisation approach to landmark points on the extracted features.

3. To develop an interactive tool to ease the visualisation of the overall feature extraction and landmarking process.

1.6 Overview of Report

The organisation of the report:

Chapter 1 provides an introduction to automatic landmarking process, and particularly to challenges and aim and objectives.

Chapter 2 gives the background of this research work and the literatures of different methods for image pre-processing and face landmarking techniques which were proposed in the recent years on existing face applications.
Chapter 3 details the methods for automatic landmarking process of this research.

Chapter 4 presents the implementations, result and analysis of the automatic landmarking process.

Chapter 5 details and concludes the report by discussing the contributions, limitations and future works for this research.
Chapter 2: Face Landmarking Techniques

2.1 Introduction

This chapter will discuss the various existing methods in the field of image processing and face landmarking techniques on 2D and 3D face data. Existing pre-processing methods on face images, which usually include techniques in illumination artifacts removal, noise removal, holes removal and geometric correction approach, will be discussed in this chapter. As the research’s aim on automatic landmarking on face image, therefore, we will focus on reviewing existing work and research on face landmarking techniques, particularly on 2D and 3D face data.

Face landmarking is defined as the detection and localisation of certain characteristic and features on a face. It is a vital process for many face analysis applications. In any typical facial applications, face landmarking plays an influential role is face expression analysis, face animation, face registration, face recognition and face tracking. In addition, face landmarking could be used for facial identification in digital images, face editing software, lip reading interpretation and so forth.

Obtaining face landmarks are a first stage towards a face data processing at a semantic level (Whitmarsh et al., 2006). The commonly used face landmarks are the corners of the eyes, the tip of the nose, the corners of the mouth, the chin, around the cheek area, and etc. Landmarks on the tip and the bridge of the nose, based on literature, have minimal changes with facial expressions. Therefore, they are relatively very stable and robust for face processing work (Celiktutan et al., 2013).

The process of landmarking, whether it is done manually or automatically, is a necessary stage in all face processing applications. Landmarks could be used to provide for
identification of different facial features on human faces. Therefore, it is an important process and the examples of the face applications are stated as follows:

i) Facial feature tracking - Application of human facial features recognition to automobile security and convenience. The system is employed to detect and track the person. Therefore, automatic face landmarking would be useful in the automation and security industries.

ii) Facial modelling and animation - Computer-aided games, which generate facial expressions, can be animated by tracking facial features on the human face. Automatic facial landmarking is able to be a main component to identify and track the user’s facial expressions.

iii) Expression analysis - The localisation of key feature points, such as the corners of the eyes and mouth, is useful to aid face recognition. The candidate facial features will be able to landmark before the face recognition process.

iv) Face recognition - An automatic identifying or verifying a person from a digital image or a video sequence. Some facial recognition algorithms identify faces by extracting landmarks, or features, from an image of the subject's face.

2.2 Quality Improvement through Pre-processing

Pre-processing of face images is an important step in many applications. The most common image acquisition technologies are light-based whereby different light sources could affect the geometry of the face model (Bowyer et al., 2006). These technologies are sensitive to specular, oily regions and regions of facial hair. It is common to obtain some artifacts such as spikes, noise and holes even under the ideal illumination settings. For instance, the impractical areas of facial hair, shoulder and hair will be captured. Therefore, there is always some pre-processing before a method engages in the face landmarking process for quality
improvement. Those typical techniques are as follows: illumination artifacts removal, noise removals, holes removal and moderate geometric corrections.

2.2.1 Illumination Artifacts Removal

Face landmarking generally meets the challenges of illumination and occlusions where some landmarks can be hidden (Szeptycki, 2012). Therefore, the captured face image which is subjected to illumination compensation can be operated locally or globally in pixel wise. One example of pixel wise normalisation is Center-Surround Divisive Normalisation (CSDN) (Meyers & Wolf, 2008). It is defined as each pixel is divided by the mean value of a block around it; another example is rank filtering where pixels around the block are ranked, and the central pixel assigned its rank value and all assignments are stretched to the [0, 255] interval.

The local normalisation technique is applied to an image which can eliminate the effect of uneven illumination meanwhile keeping the statistical properties of the processed image as in normal lighting condition. This can be achieved by using filtering with Laplacian of Gaussians or using a facet model as in (Xie & Lam, 2006). Moreover, the histogram equalisation is the example of global normalisation.

2.2.2 Noise Removal Techniques

Removal of the noise artifacts is a crucial step for quality improvement. The removal of those artifacts can be divided into two categories - filtering and geometrical analysis. The noise often occurs during the acquisition step and it is called as spikes. The frequently used technique to remove spike artifacts is the median filter (Zhao et al., 2009; Faltemier et al., 2008; Mian et al., 2007) by calculating the median value of a neighbourhood of a point, and replaced its value by the calculated median. This technique led to remove noise or some insufficient environment conditions, but do not preserve edges.
In Aminnejad et al. (2011), a set of simple techniques of pre-processing and feature extraction was proposed. The CASIA database is applied in this experiment. Median filter is applied in order to remove spike noise and holes filling of the range data. The facial depth values were scaled in between 0 to 255. The median values were replaced by the calculated median which led to remove spike noise and achieved a smooth processed range data. However, the sides of the head, forehead and ears are eliminated. The elimination from these facial areas may also cancel the potential features in the extraction process.

Another filtering technique was applied in (Wang et al., 2009; Colombo et al., 2006), where a Gaussian filter was adopted to remove the spikes and smooth the data. However, these techniques are simple and effective filter impulse noise removal, they also tend to remove the fine details and may change the entire model. As the input models differ in term of resolutions, it is necessary to control the kernel matrix size to adjust the intensity of the filtering technique on the input model.

Tyagi and Singh (2013) proposed a comparison between Gaussian filter and Gabor filter (Nixon and Aguado, 2008) for the pre-processing technique. They simulated the input images by using different threshold value settings. Gaussian filter showed a good result as compared with Gabor filter. It is able to smooth the overall shape of the image and improve some broken pints on ridges.

There is a different group of method which is relying on geometrical property analysis. In Bowyer et al. (2006) evaluated the difference of the angles between the optical axis and the point’s local surface normal. The point was considered as a spike if the angle was greater than a threshold value of 800. Another technique was proposed by Bronstein et al. (2005). The authors applied discrete gradient norm by calculating on the vertex. A potential spike was removed if the vertex contained a high value of the norm.
It is not obvious to conclude which method has the highest accuracy in the preprocessing process. The input data may differ based on different interest areas. However, the summation of the literature showed that Gaussian filter is higher ranked compared to other methods in range images. It is an effective technique to remove such artifacts.

2.2.3 Holes Removal Techniques

As it was described earlier, the second type of artifacts which often occurs is holes. Holes are mainly created based on absorption of light in the areas such as hair, eyebrows, beard or pupil. This happens because the eye is translucent and the light stripe is refracted instead of being reflected (Queirolo et al., 2010).

In range images, holes are easy to be found by examining the missing points which are surrounded by the valid neighbours. Faltemier et al. (2008) localised holes by simply interpolating x, y, z coordinates based on the surrounding points. The authors in (Zhao et al., 2009) applied morphological reconstruction to locate the holes and cubic interpolation is applied to fill the holes. This cubic interpolation method has also been used in Mian et al. (2007).

The problem of holes on range images is not complex and it can be solved easily by using simple interpolation. A various methods range from missing points localisation on range image, analysis of the number of connections between vertices and also the number of triangles joining one edge.

2.2.4 Geometric Corrections Techniques

The landmark localisation applies the knowledge of the geometric relationship, such as angles and distances between landmarks and shape characteristics. This knowledge can be adapted to a set of rules which used as a set of statistics of point-to-point angles and distances subtended by local ensembles such as the triple landmarks (Çeliktutan et al, 2013). There are
algorithms like Viola-Jones (Viola and Jones, 2001), Gabor filters (Shih and Chuang, 2004), projection histograms (Baskan et al., 2002) and so on which help to detect the eyes and the mouth. Once a few facial components are detected, the geometry information can be applied to initialise the search in a reduced search area for the remaining ones. If the certain landmarks are missing or the detected landmarks do not meet the condition, the search is reinitialised via the geometric face model. This helps to improve on time consuming during the landmark detection.

2.3 Face Landmarking Techniques

Face landmarking techniques can be categorized in various ways. It is based on the type or modality of the data, such as still image, range image, video sequence or 3D image. It is difficult to define a clear-cut distinction among the techniques since algorithms often share techniques common to more than one category (Çeliktutan et al., 2013). Commonly, it is categorized based on their type of information and methodology used (Phimoltares et al., 2007). These authors have categorized the techniques into five categories, namely geometry-based, colour-based, appearance-based, edge-based and motion-based face landmarking.

In Çeliktutan et al. (2013), the authors divided the face landmarking techniques into two basic categories: model-based methods and texture-based methods. Model-based method is also known as shape-based methods which consider the whole face image shape. The face shape is studied and labelled during the training images; it is then fitted to the proper shape of an unknown face during the test stage. Texture-based methods search each landmark or local groups of landmark independently without any guidance of the model from the training set. These two main categories are further split into two sub-categories as illustrated in Figure 1.1.
These two main landmarking approaches can be further split into two sub-categories. The model-based methods are split into implicit methods and explicit methods. Implicit methods are those algorithms using a neural network applied to the whole face; whereas explicit methods can be included under the topic of graph methods and active appearance methods. Likewise, explicit are discussed under the sub-categories of transform-based methods and template-based methods.

Hence, the face landmarking techniques can be grouped into different categories. The sub-categories of both landmarking approaches will be discussed briefly in the following sections.

2.3.1 Model-based Methods

Model-based methods classify the whole face image in order to employ the face shape to obtain landmark. Based on the sub-categories of model-based methods, explicit methods are more popular as compared to implicit methods. For instance, prime methods of explicit methods are Active Shape Model (ASM) and Active Appearance Model (AAM). There are only a few researches about implicit methods. Though, both sub-categories will be briefly listed as a review.