Feature Extraction and Localisation on 2.5D Face Range Images

Pui Suk Ting
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak, Malaysia
polly_sukting@hotmail.com

Jacey-Lynn Minoi
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak, Malaysia
jacey@fit.unimas.my

***Abstract*— In this paper, we propose a method for semi- and fully automatic landmarking detection on raw face data using feature extraction and feature localisation methods. This approach is essential in any face application such face registration, analysis and face recognition methods. This approach involves locating distinct face features, such as the corners of the eyes, the tip of the nose, the chin and etc., without human manual landmarking intervention. Automatic landmarking has a number of advantages over manual landmarking. The process of manual landmarking is time consuming, error prone and limited in accuracy. We will present the accuracy of the landmark detection based on the threshold values and the interactive tool that was also developed to give a better visualisation of the landmarking process. The threshold values are analysed and generalised based on the best detected and extracted important keypoints or/and regions of facial features. We employed the proposed approach on 2.5D range face images. The results of the automatic detection and localisation based on the extracted facial features and candidate landmarks will be shown in this paper.**

Keywords- automatic landmarking, face features, Gaussian Smoothing, Mean and Gaussian curvatures, Gaussian Pyramid

I. INTRODUCTION

The human vision could perceive edges, corners of the objects or changing of colours, without any difficulties. On a face, locating the eyes and mouth is easy for human. However, a computer vision system is unable to do such task as easily and effortlessly as a human [1]. The human vision system and the brain mechanisms that are responsible for detecting such distinct features are so complex that despite the work of neurobiologists, mathematicians and computer scientists, it is not yet possible to replicate a face features detection system accurately.

In this research, we developed a new semi- and fully automatic landmarking detection method using feature extraction and feature localisation approaches to place landmark points on palpable face features such as the tip of the nose, chin, eyes and mouth. We aim to make the landmarking process as automatic as possible on raw face image data. To the best of our knowledge, there are not many research works on automatic landmarking on acquired raw data. Automatic facial features detection and landmarking process holds a number of advantages over human manual landmarking. In large data particularly, detecting and landmarking points manually is time consuming, error prone and limited in accuracy. Automatic landmarking is still useful in many image processing applications and it is an important component for face registration, analysis and recognition methods. We will access the accuracy and the robustness of the proposed approach on 2.5D range face image data.

Comparative studies have demonstrated that automatic landmarking is a challenging problem due to the complex structure of human face, cluttered background, face variations and changing of acquisition condition. Subsequent works in the literature has demonstrate successful automatic landmarking approaches to deal with the problems.

Over the last decade, 3D images have become popular with the advancement of 3D sensor and camera technologies; alongside with 2.5D range images. Range images (or 2.5D images) different than 2D images and it has a number of added advantages over 2D and 3D images. A 2.5D image is defined as a simplified 3D (x, y, z) surface representation that contains the depth (z) value(s) for every point on a (x-, y-) plane [2]. One can think of a 2.5D image as a grey-scale image, where the black pixel corresponds to the background, while the white pixel represents the surface point that is nearest to the camera [3]. 2.5D face images enable depth perception and allow one to manipulate the image alike a 3D image. In addition to range data, colour perception on a face image is also possible. The information from the range image can be extracted to derive different features regions. Therefore, 2.5D range images is used as a dataset to define the keypoint descriptor by extracting the facial surface information.

We present a method for feature extraction and localisation in range images. Our goal is to develop an automatic landmarking process in cluttered images with different facial variants. In initial attempts to compare the accuracy of the landmark points, threshold values were analysed based on the results of the feature extraction. This stage is known as the semi-automatic landmarking process. Our method follows a coarse-to-fine strategy process for the localisation of facial landmarks. There are three main stages before localising the landmark points. Gaussian Smoothing is applied in the smoothing phase to extract the weighted regions by setting the upper and lower threshold values. Scale-space is then constructed by taking the Difference of Gaussian (DoG) images at different the scales/region of interest. Within the scale-space, mean-weighted of the curvature values of the elements is computed to estimate the centres. These HK regions are curvatures map and they are colour coded to ease visualisation. Lastly, Gaussian pyramiding is applied. We used the '*Reduce*' and '*Expand*' methods to acquire the filtered regions and to localise the candidate landmarks and/or best regions. We have also developed an interactive Graphical User Interface (GUI) to visualise the landmarking process by manipulation the extracted facial features.

The outline of this paper is as follows: In Section II, we describe the previous works related with our research. In Section III, we explain our automatic landmarking method. In Section IV, the result and discussion are presented. Finally, Section V offers conclusions and directions for future research.

II. Related Works

Facial landmarking with automatic system has aroused interest in the increasing development of face applications in recent years. It has the potential to be very beneficial ranging from face registration to facial expression recognition [4]. Initial candidate facial landmarks selection and landmark registration steps are necessary in almost any application when processing of facial data. Therefore, registration based on facial landmarks correspondence is the most crucial step to make a system fully automatic [5].

There are a number of publications whereby numerous face researchers have experimented different techniques and methods to detect and select candidate facial landmarks. These works are categorised by prior knowledge using facial geometry-based analysis, curvatures analysis and so on [4].

Nowadays, the most common and popular method for face images analysis and processing relies on curvatures on their facial features [6]. This approach has been widely used since the 1980s introduced Besl and Jain [7] using Mean (H) and Gaussian (K) curvatures to segment facial surface into eight different surface types. In the early works, these researchers [8, 9, 10, 11, 12] adopted H&K curvatures method and experimented it on range images and then on human faces to extract features based on shape and curvature of the face surfaces.

Koenderick and van Doom [13] introduced Shape Index (SI) method by decoupling the extracted shape and magnitude of the curvedness. In [14, 15, 16], the positions of the corners of the eyes and mouth, the nose and the tip of the chin are extracted. They have also developed a heuristic method to identify the tip of the nose more efficiently by using statistical model and filtering to extract the candidate landmark positions.

Dibeklioglu et al. [17] developed an automatic 3D facial landmarking algorithm by using statistical and heuristic approaches. They used statistical landmark localisation based on an analysis of local features in order to detect the each landmark separately. A heuristic method (a curvature-based method) is also introduced to localise the tip of the nose under severe conditions. This algorithm can be employed to model any facial landmarks but data with pose variations between testing and training sets were not taken into consideration.

Later, Nair and Cavallaro [18] presented a method to detect candidate facial landmarks on 2.5D scans. They applied the shape index and the curvedness index to extract feature points, the tip of the nose and the corners of the eyes. The feature points are fitted to the dataset according to three selected control points for the registration. However, the method is not applicable to pose, self-occlusion and missing of data.

In [19, 20], the authors presented methods for better detecting facial landmarks on 2.5D images. The candidate landmarks are the inner and outer corners of the eyes, the corners of the mouth, the tip of the nose and chin. In order to locate the candidate landmarks, the shape index, extrusion maps and spin images were used. The candidate landmarks are identified using statistical facial landmark model.

III. Method

The proposed method consists of three stages. In the first stage, Gaussian Smoothing is applied on the acquired raw range images to eliminate lower contrast of the keypoints. Secondly, H&K curvature is used to determine the informative primitive surfaces. Lastly, the Gaussian Pyramid is constructed to perform "*Reduce*" and "*Expand*" operations [21]. Localising landmark points is then computed on the extracted facial features and/or regions. Figure 1. illustrates the overview process of the automatic landmarking stages.

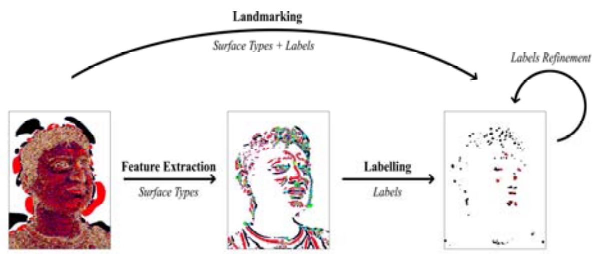

Figure 1. An overview of automatic landmarking process

A. Gaussian Smoothing

In the initial stage, Gaussian Smoothing is applied to construct the scale-space and to detect the local maxima and minima of candidate keypoints. Interpolation of nearby pixels from the calculated DoG images is used to determine the positions of each candidate keypoint. Keypoints with low contrast are removed and the responses along edges are eliminated. Each pixel is examined by comparing its eight neighbours at the same scale in a DoG image and the nine corresponding neighbours at the neighbouring scales. A candidate keypoint is selected if the pixel is a local maximum or minimum. The properties of the keypoint are measured to the keypoint orientation, which provides rotation invariance.

B. Mean (H) and Gaussian (K) Curvatures

A 2.5D range image contains *z*-depth on each pixel point. The depth value of each pixel is represented by grey intensity and the nearer it is to the camera, the brighter the grey values. This depth information could also represented in any RGB colours.

HK curvature is scale/resolution and orientation invariant, whereby spaces are constructed in order to classify primitive surfaces (patches) into types [22]. There are nine types of patches such as peak, ridge, saddle ridge, flat, minimal, pit, valley, saddle valley and undefined.

In this process, the primitive surfaces are examined to generalise threshold values by bootstrapping and resampling DoG images.

C. Gaussian Pyramid

Gaussian pyramiding is a method to down-sample the image into half its size. The "*Reduce*" operation will also be applied during Gaussian pyramiding. The elements of the image are halved in both resolution and successive scales. On each scale, H&K maps are computed. A new pyramid of HK maps is obtained and used at every scale. Figure 2 shows the higher to smaller levels of the pyramiding operation. The "*Expand*" operation is computed to widen the primitive surface labels at each level of H&K pyramid to the original size.

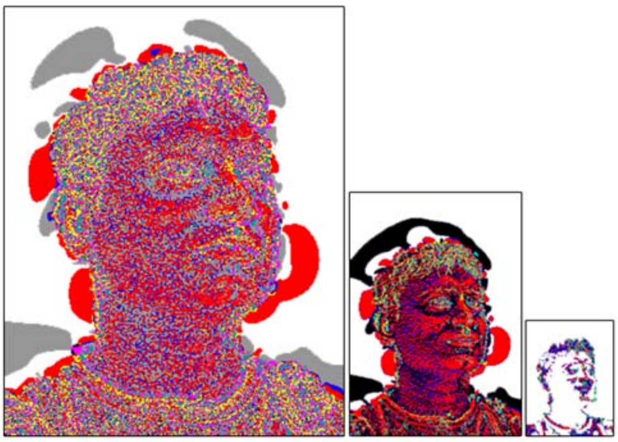

Figure 2. Gaussian Pyramid "Reduce" operation.

The original size of the range image is 1944 pixels by 1296 pixels and at this scale, representative features are difficult to detect as the support region. There are three levels used by pyramiding on the range image to obtain the valuable information. The size of the pyramid level in the Gaussian Pyramid "*Reduce*" and "*Expand*" is as follows:

TABLE 1. SIZE OF THE IMAGE AT EACH PYRAMID LEVEL

Level	Size (pixels)
1	1944 x 1296
2	972 x 648
3	486 x 324
4	243 x 162

D. Localisation

The next step is to compute and put landmarks on the extracted primitive surfaces selected from the HK maps. It computes and places landmarks on the extracted primitive surfaces from the H&K maps after Otsu's method. Otsu's method [23] calculates a global threshold value by calculating the spread within each of the classes. The aim is to minimize the weighted within-class variance and maximize the between-class variance of the thresholded black and white. The details of this algorithm can be found in [23]. The primitive surfaces on HK maps are converted into grayscale HK maps. For example, if the user selects 'Peak' surface, the grayscale color would be = R150 G150 B150 and the background would be black in color = R0 G0 B0.

We have also set the minimum threshold size of the local pixel area of the primitive surface for landmarking. The threshold size can be adjusted through the developed interface tool.

The identified landmarks are labeled. If the generated landmarks are undetectable, the user may alter the landmark threshold size. If the landmark label is missing, the user needs to reduce the landmark size.

IV. RESULT

In order to make the landmarking process as automatic as possible on raw face image data, we developed interactive tool is to ease the visualisation of the features extraction and localisation processes and to improved the processing time and the performance on the labelling of the candidate keypoints. All the nine keypoint descriptors being extracted, and each of them is examined by a set of minimum local pixel sizes for landmark labels.

We applied the experiments on three different sets of range images. Each of the keypoint descriptors is examined by a set of minimum local pixel sizes for landmark labels. Only four keypoints are sufficient on face data to be extracted and the descriptors are peak, ridge, valley and flat (see Figure 3). Each of the extracted features image is processed with binary conversion in order to execute the landmarking process. One of the results of the extracted primitive surfaces are shown in Figure 4(a). We then execute the automatic landmarking process on these surfaces. Threshold values ranging from 10 to 70 pixel sizes were tested before it could be classified as a generalized threshold value. After the labelling process is completed, the landmarks are computed and placed on the face images. These values are tested accordingly and repeatedly until the highest accuracy of landmarks is obtained. The default threshold size selected is 40 pixels. This value is chosen as it has shown a sufficient average area in the localisation stage.

From the semi-automatic landmarking experiments, we have analysed and identified that the threshold values from 30 to 60 pixel size is sufficient to extract landmarks. Numerous labels were detected on the face images within this range. The computed landmarks are manually counted only on the region of the face, whereas the hair and shirt areas are eliminated as noise.

Figure 3. The extracted H&K features and converted to binary images.

Figure 4(c) shows the successful landmarks on specific areas of interest such as the tip of the nose, the bridge of the nose, the eye and so on. There are numerous incorrect landmark labels located not on the face areas especially when the threshold values were set below 30 pixels. Threshold value set to 60 pixels or more has only three ore less landmark labels detected. Hence, we deduce that the threshold pixel values should range in between 30 to 60 pixels size for this set of face range data sets. The results of the pixel size values to landmarks are summarized in Table 2.

The results of the experiment have shown that we able to label landmarks on peak, ridge, flat, and valley. The three identified primitive surfaces - peak, ridge and flat – are the most robust and reliable features to identify landmarks for our dataset. The commonly identified face landmarks are the tip of the nose, the corners of the mouth and the chin. For instance, the primitive surface of peak has successfully labelled the tip of the nose automatically. On the contrary, ridge and flat primitive surfaces mainly identify features on the hair, shoulder and the shirts. These features regions are indicated as noise and needs to be eliminated from the candidate labels. There are plenty of flat features landmarked on the hair. The valley primitive surface was detected as a landmark label but did not show a stable result in the three range images. No valley features found in the face range image set 1, and only one landmark label on valley feature is detected in range image set 2 and 3 and both are located on the cheek of the face. From this result, we deduce that valley primitive surface might not be beneficial for landmarking our dataset. The other five primitive surfaces - saddle ridge, undefined, minimal surfaces, saddle valley and pit - do not provide any landmark labels. Therefore, those primitive surfaces can be concluded as insufficient for landmarking candidate in this experiment. Even though, valley feature is extracted, it still outperformed the other five primitive surfaces and therefore it is still less robust as compared to peak, ridge and flat surfaces. We have concluded that the pixels after these feature extractions are not sufficient to be used for landmarking detection and localisation labeling.

TABLE 2. THE SUMMARY OF THE NUMBER OF LANDMARKS CALCULATED ON THE MOST USEFUL SELECTED PRIMITIVE SURFACES DURING THE LANDMARKING PROCESS

Primitive surfaces (landmark labelled)	Threshold value (pixel size)			
	30	40	50	60
Range image set 1:	Number of landmark labels			
Peak	5	3	2	None
Ridge	3	3	2	2
Flat	7	6	3	2
Valley	None	None	None	None
Range image set 2:	Number of landmark labels			
Peak	4	3	2	1
Ridge	8	7	5	3
Flat	8	8	7	7
Valley	1	1	1	None
Range image set 2:	Number of landmark labels			
Peak	4	3	2	1
Ridge	5	4	2	2
Flat	4	4	2	2
Valley	1	1	None	None

V. ANALYSIS AND DISCUSSION

Feature extraction and localisation of the landmark processes were performed on three available face range images corresponding to three different persons. The results have shown that two to eight facial landmark labels are identified correctly on each range face image. This is verified visually. The best result obtained on all the range images could identify landmark automatically on the tip of the nose. The other identified facial landmarks are at the corners of the eyes, the corners of the mouth and the chin.

The automatic landmarking technique was compared on three sets of manual landmarking coordinates on the same dataset by three experts. The coordinates of the tip of the nose and the chin were registered to perform an accuracy test on the automatic method. The Root Mean Squared Error (the difference) was calculated between the data points in the automatic method with each manual registration points.

$$Accuracy, A = 2.4477x0.5\times (RMSE\text{x} + RMSE\text{y})$$

Based on the calculation, the automatic method performs relatively well for the detecting the tip of the nose and the chin, with low error and high accuracy. The accuracy is obtained by direct comparison of pixels with the automatic and manual landmark points. The average accuracy pixel differences for the tip of the nose is by 0.956 pixel while the chin is by 1.739 pixel.

It can always be expected that there is a slight variation when comparing the manual landmarking results, either due to user manual selection/mouse positioning and/or selection of landmarks [24]. However, the pixel differences that were found between the automatic and manual landmarking are negligible.

On the other hand, we have also identified limitations occurred during the experiments. Currently, there are only three range images being tested and the file format is fixed. Different file formats should be applied and tested in order to improve the robustness of the current landmarking system. A method could also be used to eliminate noise such as hair, shoulder and shirt. In current work, a pre-processing method is applied to delete those regions. A better quality and pre-processed image shall be tested to examine the performance of the current system.

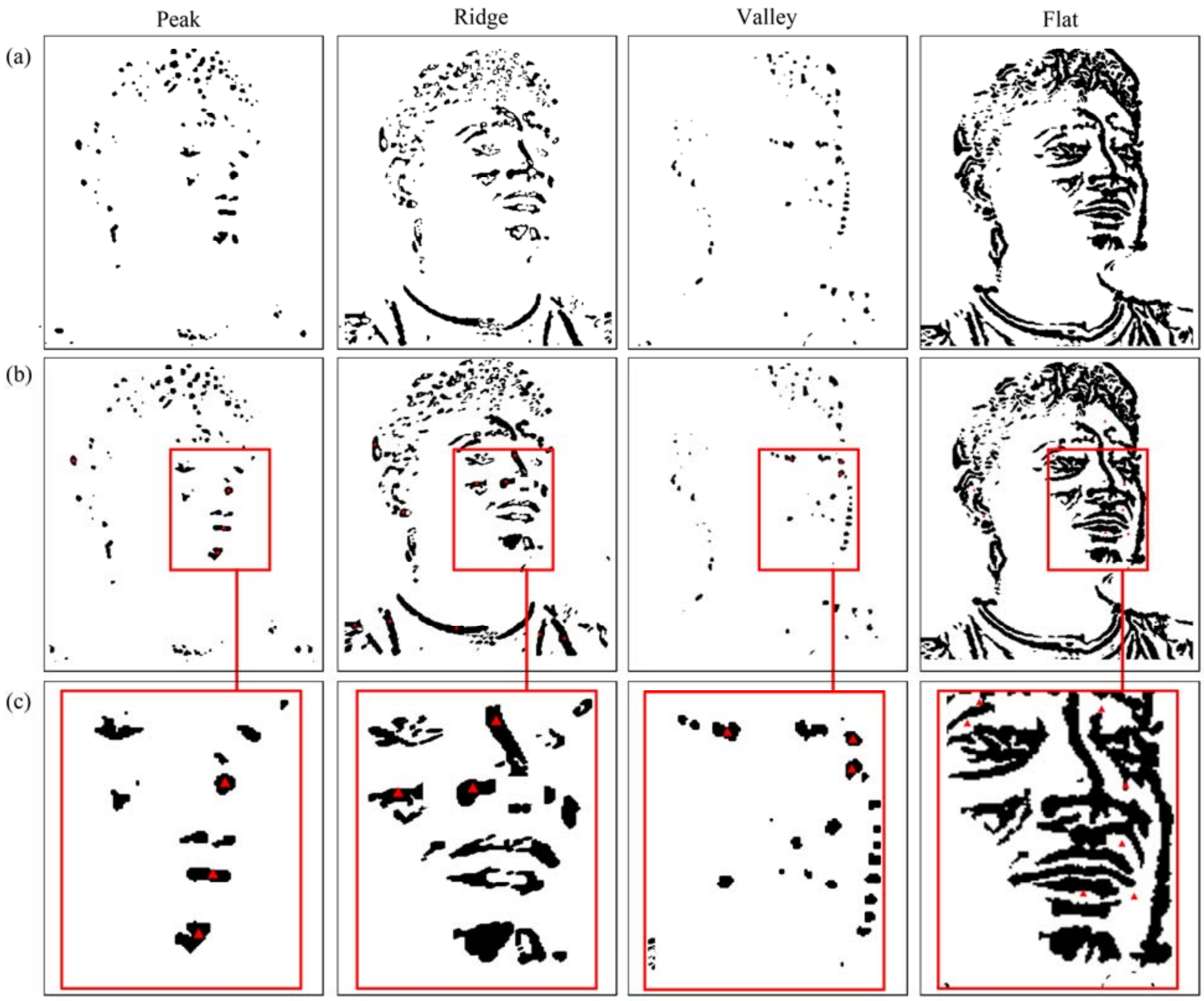

Figure 4 (a) Extracted primitive surfaces; (b) labeled landmarks and (c) enlarged landmarks images.

VI. CONCLUSION

In conclusion, we have successfully implemented a semi-automatic landmarking method. The interactive tool helps to execute features extraction and landmarking process in a better visual manner. We tested the method on a set of different threshold values of local pixel sizes. The results have shown that after executing the method with different threshold values, the identified output of the landmarks is similar. Therefore, we could label the landmark automatically.

We have also successfully developed a feature localisation stage by applying Otsu's method and geometry centroid approach. The details of the work can be found in [25]. This newer framework helps to landmark points automatically on the extracted features. From the extracted primitive surface types, binarisation approach was taken to determine the informative regions, which would be suitable for the landmarking of points. Geometry centroid approach is further applied to calculate the landmark point. The small regions were eliminated as noise whereby the informative regions were labelled according to the threshold values set.

VII. LIMITATIONS AND FUTURE WORKS

The limitations of the projects are as follows. Firstly, the numbers of the dataset are limited for the experiment. There were only three .dat range images being tested. And also the file format is fixed in the current system. Therefore, the accuracy of the landmark labels is very much based on a small set of range images. A different set of face shapes, ethnic groups, gender and age group with various file formats could be tested in the future work. These images might be associated with different sets of landmark labels. The number and nature of the landmarks might also be varied from the current results. The frontal images could be tested to improve the detection and location of facial features by using a geometric model of the face and exploiting features only on the face region. The three main points such as the eyes corners, the tip of nose and the chin by generating a face template will help specifically in face registration. Another improvement that can be done on the range images is the pre-processing approach to extract only face regions. Regions containing hair, shoulder and shirt should be removed.

This project contributes to various researches such as face localisation, faces landmarks extraction, face analysis and face expression detection and not limited to facial feature extraction. In the future, we will also conduct experiments on general object detection.

ACKNOWLEDGMENT

This research is funded by Science Fund (e-Science) grant number 01-01-09-SF0097. Thanks to UNIMAS and T.C. for the 2.5D image.

REFERENCES

[1] O. Çeliktutan, S. Ulukaya and B. Sankur, "A comparative study of face landmarking techniques," EURASIP Journal on Image and Video Processing, no. 1, 2013.

[2] V. Tag, "Multimedia Making It Works," Mc Graw Hill, 8th ed., 2011, pp. 142.

[3] A. F. Abate, M. Nappi, D. Riccio, G. Sabatino, "2D and 3D face recognition: A survey," Pattern recognition Letters, 2007, vol.28, no. 14, pp. 18885-1905.

[4] P. Szeptycki, M. Ardabilian and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," in Biometrics: Theory, Applications, and Systems (BTAS 09), IEEE 3rd International Conference, 2009, pp. 1-6.

[5] P. Perakis, G. Theoharis and I. A. Kakadiaris, "3D facial landmark detection and face registration," Technical report, University of Athens, Greece, 2010.

[6] P. Szeptycki, "Processing and analysis of 2.5D face models for non-rigid mapping based face recognition using differential geometry tools," PhD thesis, Ecole Centrale de Lyon, 2012.

[7] P. J. Besl and R. C. Jain, Gupta, "Three-dimensional object recognition," ACM Computing Surveys (CSUR), 1985, no. 1, pp 75-145.

[8] E. Trucco, and R. B. Fisher, " Experiments in curvature-based segmentation of range data," in Pattern Analysis and Machine Intelligence, IEEE Transactions, 1995, no. 17(2), pp. 177-182.

[9] G. G. Gordon, "Face recognition based on depth and curvature features," in Computer Vision and Pattern Recognition, Proceedings CVPR 92, IEEE Computer Society Conference, June 1992, pp. 808-810.

[10] H. T. Tanaka, M. Ikeda and H. Chiaki, "Curvature-based face surface recognition using spherical correlation, principal directions for curved object recognition," in Automatic Face and Gesture Recognition, Proceedings, Third IEEE Internation Conference, April 1998, pp. 372-377.

[11] T. K. Kim, S. C. Kee and S. R. Kim, "Real-time normalization and feature extraction of 3D face data using curvature characteristics," in Robot and Human Interactive Communication, Proceedings, 10th IEEE Internation Workshop, 2001, pp. 74-79.

[12] A. B. Moreno, A. Sánchez, J. F. Vélez and F. J. Díaz, "Face recognition using 3D surface-extracted descriptors," in Irish Machine Vision and Image Processing Conference, September 2003, vol. 2003.

[13] J. J. Koenderink and A. J. van Doom, "Surface shape and curvature scales," image and vision computing," 1992, no. 10(8), pp. 557-564.

[14] D. Colbry, G. Stockman and A. Jain, "Detection of anchor points for 3D face verification," in Computer Vision and Pattern Recognition, Proceedings, IEEE Computer Society Conference, June 2005, pp. 118.

[15] X. Ju, A. Jain and D. Colbry, "Matching 2.5D face scans to 3D models," in Pattern Analysis and Machine Intelligence, IEEE Transactions, 2006, no. 28(1), pp. 31-43.

[16] D. Colbry, "Human face verification by robust 3D surface alignment," PhD thesis, Michigan State University, 2006..

[17] H. Dibeklioglu, A. A. Salah and L. Akarun, "3D facial landmarking under expression, pose, and occlusion variations," in Biometrics: Theory, Applications and Systems, BTAS 08, 2nd IEEE International Conference, September 2008, pp. 1-6.

[18] P. Nair and A.Cavalloro, "3D face detection, landmark localization, and registration using a point distribution model,"in Multimedia 11, IEEE Transaction, 4 June 2009, pp. 611-623.

[19] P. Perakis, G. Passalis, T. Theoharis, G. Toderici and I. Kakadiaris, "Partial Matchin of interpose 3D facial data for face recognition," in Biometrics: Theory, Applications and Systems, 3rd IEEE Internation Conference, Arlington, VA, 28-30 September 2009, pp. 439-446.

[20] P. Perakis, T. Theoharis, G. Passalis and I. Kakadiaris, "Automaic 3D facial region retrieval fom multipose facial datasets," in 3D Object

Retrieval, Proceedings Eurographics Workshop, Munich Germany, April 2009, pp. 37-44.

[21] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, 2004.

[22] H. Guo, K. Zhang and Q. Jia, "2.5D SIFT Descriptor for Facial Feature Extraction," in Interlligent Information Hiding and Multimedia Signal Processing, Sixth International Conference, 2010.

[23] N. Otsu. A threshold selection method from gray-level histograms. In Automatica, 11(285-296), 23-27, 1975.

[24] J. Shi, A. Samal, and D. Marx, " How effective are landmarks and their geometry for face recognition?" in Computer Vision and Image Understanding, 2006, no. 102(2), pp 117-133.

[25] S. T. Pui and J.L. Minoi, "Feature Extraction and Localisation using Scale-Invariant Feature transform on 2.5D image,' in Computer Graphics, Visualization and Computer Vision, 22nd International Conference, June 2014.