# 3D Face Recognition Using Multiple Features for Local Depth Information

**Fatimah Khalid, Lili N. A.**

Department of Multimedia, Faculty of Computer Science and Information Technology
University Putra Malaysia, 43400 UPM Serdang,Selangor Darul Ehsan

**Summary**
In this paper, we recognized multiple features from the local depth information of distance and angle calculation. These features are calculated from the twelve salient points by considering the distance and angle calculation. Then, fifty three non-independent features are extracted and the discriminating power is used for analyzing these features. The result shows an improvement compared to the previous work.

***Key words:***
*3D face recognition, feature localization, feature extraction*

## 1 Introduction

Face recognition is based on the computer identification of unknown face images by comparison with a database of known images. Several disciplines involved such as image processing, pattern recognition, computer vision, computer graphics, and machine learning. Since 1975, researchers in psychophysics, neural sciences and engineering, image processing, analysis, and computer vision have been investigating a number of issues related to face recognition by humans and machines.

The vast majority of face recognition research has focused on the use of two-dimensional intensity (Bowyer et al. 2004). Over a decade ago, a new research paradigm for face recognition focused on three-dimensional (3D) images, either by itself or in combination with two-dimensional (2D) images. The use of multiple imaging modalities, such as 3D and 2D images of the face refers to *multi-modal biometrics.* Recently, the curiosity interest in model-based 3D automatic recognition systems has been increased (Lee & Ranganath 2003; Huang et al. 2003). The process of 3D face recognition is the same as 2D face recognition where it involves detection as well as representation and matching.

Gupta et al. (2007) categorized 3D face recognition techniques into two main categories which is (a) based on the appearance of facial range images (appearance based) and (b) based on the geometric properties of 'free form' facial surfaces ('free form' based). They said that appearance based is also known as statistical learning based where the techniques are all in statistical approach (as shown in Figure 1). These appearances of facial range approaches are similar to 2D holistic appearance based techniques where the only difference is the employment of range images instead of intensity images. Meanwhile, ensemble approaches consisting of combinations of multiple 'free form' techniques.

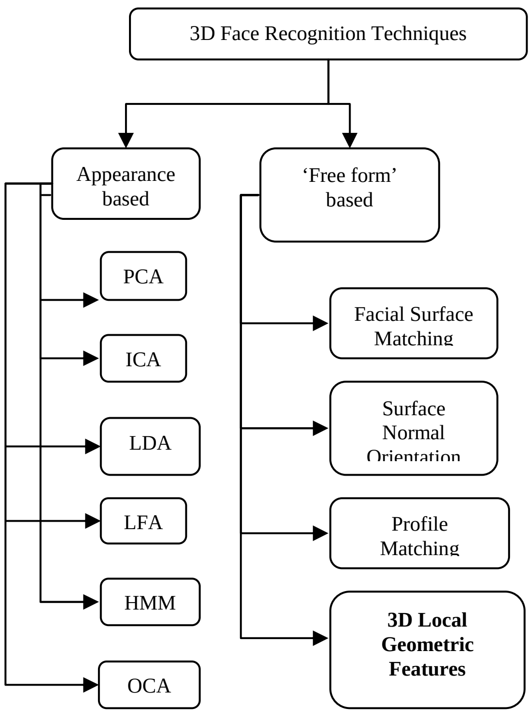


Figure 1. 3D face recognition categorization
Source: Gupta et al.

Studies considering in local geometric feature of 3D facial surfaces still few in numbers. This is due to the requirement of automatic segmentation of facial landmarks. The local shapes of facial landmarks/regions on fiducial points have been quantified. For instance, Gordon (1992) and Moreno et al. (2005) used Gaussian curvature values for segmenting the regions and lines of

Manuscript received January 5, 2009
Manuscript revised January 20, 2009

interest. Meanwhile, Xu et al. (2004) applied Gaussian-Hermite moments technique for feature extraction. Chua et al. (2000) and Wang et al. (2002) employed 'point signatures' technique for registration where it helps to find the transformation which optimally aligns the rigid regions of the two face surfaces. Wang and Chua (2005) utilized 3D Gabor filters to extract expression-invariant features and also view-invariant features by rotation-invariant 3D spherical Gabor filter.

Based on the scenario above, it reveals that in order to calculate and understand the local curvature, the researchers need to really understand mathematical calculation. In addition, as data sets become larger, the algorithms will become more sophisticated even if the reported identification rates are not as high as in some earlier works (Bowyer et al. 2006). Because of this reason, a technique has been built which does not need to understand the mathematical calculation which directly calculates the syntax of the 3D points with curvature information.

On the other hand, the points detected are not suitable for variation of pose and expression. For example, middle mouth will affect the face while smiling and this point cannot be used for expression variation. For this reason, a set of quality points in different expression and position need to be explored. For recognition, Moreno et al. (2005) used thirty local geometrical features from a set of 86 extracted from 3D human face surfaces to model the face for recognition. From their research, they found out that features such as angles and distances offer a better recognition results. The area based features resulted in the poorest discriminating power, due to their corresponding larger variation in the within-class variance of the feature according to Fisher's criterion. The recognition performance result shows that the highest is 78% for the first image and 84% for second. The result is unpromising and different features should be added to get more accuracy.

This paper discussed the feature presentation for recognition. It is the second stage in face recognition after face detection where comparing between faces based solely on a set of scalar features. It determines the calculation of the distances and angles. A vector in feature space is used to associate each face in a set of features. The feature calculation begins with the 3D data which is in text file and at the same time a simple mathematical formula is applied. This method is adopted on the basis of a small number of feature measurements. As a result, less memory is required for storage and the computationally is very simple. Although the feature calculation is obviously varied from other researcher where they based on the curvature, we also used the curvature information to detect points from project into graph. These features are calculated primarily from the set of anchor point's localization which will be discussed in the next section.

## 2 Salient point's localization

In this research, the first stage of the feature extraction from a face is to mark the salient events on the face surface in terms of points.

### 2.1 Nose tip point localization

The first step in 3D face detection is to detect the nose tip which can be easily identified by finding the highest $z$ value in the 3D face model. This rule can be used for the frontal view of a face. Furthermore, the nose tip is a distinctive point of the human face and also insensitive to any changes of the facial expression (Lu 2006). To determine whether the faces are in the frontal view, the direction of Z axis should be on top of the nose.

### 2.2 Upper nose and upper face point localization

Given the estimated nose tip, the next step is to obtain the eye and mouth region. This can be done by constraining the particular points by fixing the $X$ axis as explained in Table 1. From the plotted $YZ$ axis, mouth region and eye region can be determined. Thus, the last point at the traversed north is determined as upper face point. Besides, upper nose point can also be obtained by encountering the lowest $z$ point value.

Table 1: The upper nose and upper face anchor point detection algorithm

1. Nose tip estimated
2. Calculate the constrained eye and mouth region by fixing the $X$ axis and taking the larger and lower $Y$ axis within 0.5 neighborhoods to get the horizontal line
3. From nose tip point, points are traversed north towards the forehead until the first lowest $z$ value is encountered. This point is in the middle of inner left and right eye which called as Upper nose point.
4. Then the last point of this eye region is encountered as Upper face.

### 2.3 Chin point localization

From the vertical line segmentation, mouth region is from nose tip point until the lowest $Y$ value. Then to obtain the chin point, starting from the minimum of *Y* value, traversed along the *Y* axis until reaching the first peak point called Chin point.

### 2.4 Nose base and outer face points localization

Referred to the estimated nose tip again, the search region for nose base point location is constrained. The algorithm was explained in Table 2. These two points will be used for estimating the nose width and depth in feature extraction.

Table 2: The nose base and outer face anchor point detection algorithm

1. Nose tip estimated
2. Calculate the constrained left and right nose region by fixing the *Y* axis and taking the larger and lower *X* axis within 0.5 neighborhoods to get the horizontal line.
    a. lower – Right nose region
    b. larger – Left nose region
3. Sort ascending the left nose region points by *X* axis.
4. From nose tip point, points are traversed along *Z* axis points until get the point that is more than previous point or within 0.9 differences (the valley) called left nose base point.
5. The last point of this left nose region is encountered as left outer face point.
6. For the right nose point detection, sort descending the right nose region points by *X* axis.
7. Same as left nose base point, traversed along *Z* axis points until get the point that is more than or within 0.9 differences (the valley) with the previous point. This is encountered as the right nose base point.
8. The last point of this right nose region is encountered

### 2.5 Outer and inner eyes corner localization

Lastly, is to identify both points for eyes corner. Starting from the upper nose point, the eye corners are then determined with the same concept as obtaining the nose base points which was illustrated in Table 3. The inner and outer eye corners are obtained from two steps. First is by traversing from upper nose point until the difference of a point is more than or within 0.9 from the previous point.

Table 3: The eyes corner point detection algorithm

1. Upper nose point estimated
2. Calculate the left and right eyes region by fixing the Z axis and taking the larger and lower X axis within 0.5 neighborhoods to get the horizontal line as shown in Figure 3.19(b) which
    - Lower – right eye region
    - Higher – left eye region
3. Traversing from upper nose until get the point that is more than or within 0.9 differences (the valley) with the previous point. This point is encountered as inner eye points (inner left and right eye point).
4. Then, traversing from outer points (left and right) until reaching first peak and this will be determined as outer eye points.

Figure 2 shows all the twelve anchor points to be localized.

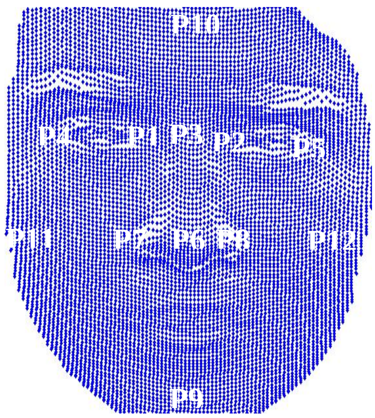


Figure 2. Anchor point locations. P4 – Right Outer Eye; P1 – Right Inner Eye; P5 – Left Outer Eye; P2 – Left Inner Eye; P3 – Upper nose point; P6 – Nose Tip; P7 – Right Nose Base; P8 – Left Nose Base; P11 – Right Outer Face; P12 – Left Outer Face; P9 – Chin; P10 – Upper Face

## 3 Face feature extraction technique

Human face is made up of eyes, nose, mouth, chin and etc. The differences between faces are shape, size and structure of the organs. So the faces are differ in thousands ways and can be described by using the shape and structure of the organs just to recognize them. Here, we extracted the distances and angles from those salient points as illustrated in Figure 2. Fifty three non-independent features were extracted from the salient point as illustrated in Figure 2. Table 4 describes how these features were obtained from the distances and angles calculation. Each feature has an associate a ranking position and identification number. All the 53 features have their own associate identification number and the order of the discriminating power where the higher discriminating power has a lower rank position while the lower discriminating power has a higher rank position. Discriminating power estimation of each feature $\Phi$ was computed using Fisher coefficient, which represents the ratio of between-class variance to within-class variance, according to the formula

$$\frac{\sum_{i=1}^{c}(m_i - m)^2}{\sum_{i=1}^{c}\frac{1}{n_i}\sum_{x\in\Phi}(x - m_i)^2} \qquad (1)$$

where $c$ is the number of classes or subjects, $\Phi$ is the set of feature values for class $i$, $n_i$ is the size of $\Phi_i$, $m_i$ is the mean of $\Phi_i$, and $m$ is the global mean of the feature over all classes (Moreno et al. 2003).

Table 4: Ordered from the best discriminating between 60 subjects to worst.

| **Rank pos** | **Id number** | **Feature description** |
|---|---|---|
| 1 | S5 | dist(P1,P2) |
| 2 | M32 | ½[dist(P1,P7)/dist(P2,P8)] |
| 3 | R26 | ½[dist(P1,P9)/dist(P2,P9)]/dist(P3, P9) |
| 4 | M30 | ½[dist(P1,P6)/dist(P2,P6)] |
| 5 | D8 | dist(P1,P6) |
| 6 | R27 | ½[dist(P4,P9)/dist(P5,P9)]/dist(P3, P9) |
| 7 | A1 | ang(P1,P3,P2) |
| 8 | M29 | ½[dist(P4,P3)/dist(P5,P3)] |
| 9 | D11 | dist(P2,P8) |
| 10 | M31 | ½[dist(P4,P5)/dist(P5,P6)] |
| 11 | A4 | ang(P2,P3,P6) |
| 12 | D3 | dist(P4,P3) |
| 13 | M33 | ½[dist(P4,P7)/dist(P5,P8)] |
| 14 | D4 | dist(P5,P3) |
| 15 | A6 | ang(P1,P2,P6) |
| 16 | M34 | ½[dist(P1,P9)/dist(P2,P9)] |
| 17 | A2 | ang(P1,P3,P6) |
| 18 | S1 | dist(P9, P10) |
| 19 | D18 | dist(P4,P9) |
| 20 | D7 | dist(P1,P7) |
| 21 | D15 | dist(P5,P9) |
| 22 | D6 | dist(P9,P3) |
| 23 | A9 | ½[ang(P4,P3,P6)/ang(P5,P3,P6)] |
| 24 | D1 | dist(P1,P3) |
| 25 | S7 | ½[dist(P6,P7)+ dist(P6,P8)] |
| 26 | D9 | dist(P4,P7) |
| 27 | D10 | dist(P4,P6) |
| 28 | D12 | dist(P2,P6) |
| 29 | A5 | ang(P5,P3,P6) |
| 30 | D13 | dist(P5,P8) |
| 31 | S8 | dist(P3,P6) |
| 32 | M28 | ½[dist(P1,P3)/dist(P2,P3)] |
| 33 | R22 | dist(P1,P2)/dist(P3,P6) |
| 34 | S9 | dist(P7,P8) |
| 35 | S3 | dist(P2,P5) |
| 36 | R23 | dist(P4,P5)/dist(P3,P9) |
| 37 | S2 | dist(P11,P12) |
| 38 | S4 | dist(P1,P4) |
| 39 | D2 | dist(P2,P3) |
| 40 | D17 | dist(P1,P9) |
| 41 | D16 | dist(P2,P9) |
| 42 | A8 | ½[ang(P1,P3,P6)/ang(P2,P3,P6)] |
| 43 | A7 | ang(P4,P6,P5) |
| 44 | R25 | dist(P4,P5)/dist(P3,P6) |
| 45 | D21 | dist(P8,P9) |
| 46 | A3 | ang(P4,P3,P6) |
| 47 | D19 | dist(P7,P9) |
| 48 | M35 | ½[dist(P4,P9)/dist(P5,P9)] |
| 49 | S6 | dist(P4,P5) |
| 50 | D5 | dist(P6,P3) |
| 51 | D14 | dist(P5,P6) |
| 52 | R24 | dist(P1,P2)/dist(P3,P9) |
| 53 | D20 | dist(P6,P9) |

The results obtained in Table 4 were used to plot the following line-graph as shown in Figure 3. Result shows getting a gradually decreased as the discriminating power smaller.

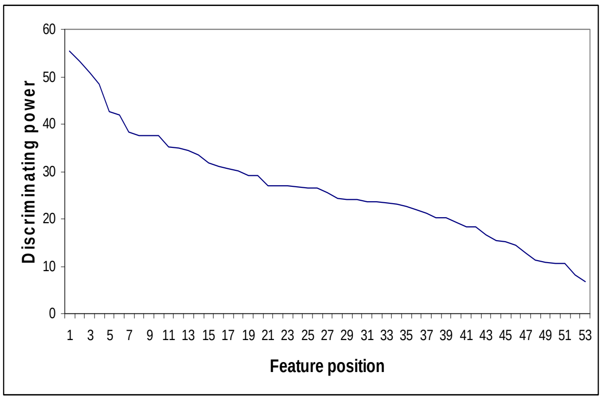

Figure 3. Discriminating power values corresponding to each feature rank position in the ordered list according to the discriminating powers.

## 4 Recognition Rates

The experiments were conducted and evaluated based on several conditions from seven different sets of faces where each person has seven different conditions: looking up, looking down, two frontal view, frontal gesture, frontal laugh and frontal smile. The test image will be the set that is not used for training. This section also reports the best result, some observations and discussion on the kind of conditions that were observed. Six images were used for training and the rest will be used for testing. Each conditions each test image. Then, the results obtained will be presented and analyzed individually. Figure 4 revealed the result of the experiments and showed that both frontal views were the highest matching rate. Meanwhile, the detailed result was illustrated in Table 5.

From the result reported, condition of looking up data set is the lowest percentage among all the data set. Meanwhile, as been expected, both frontal data sets acquired the best result. This is followed by the frontal smile data set, frontal gesture data set and the frontal laugh and looking down data set.

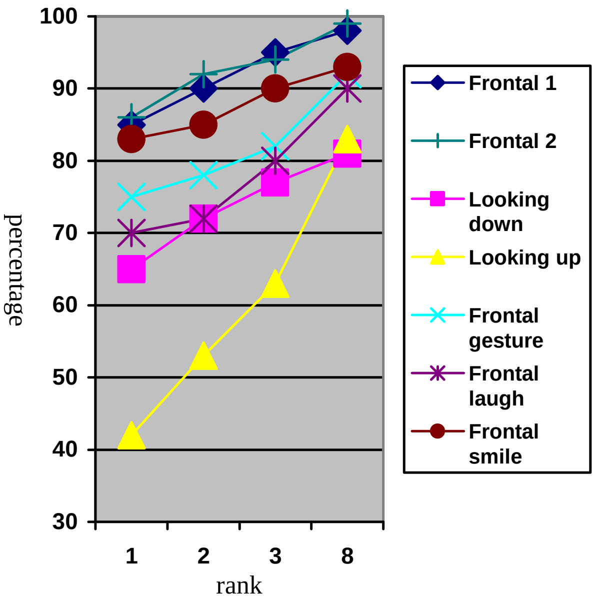


Figure 4 Results of 7 Matching Experiments

Table 5 Results of 7 conditions in percentage

| **Data set** | **Rank matched** | | | |
|---|---|---|---|---|
| | **1** | **2** | **3** | **8** |
| **Frontal 1** | 85 | 90 | 95 | 98 |
| **Frontal 2** | 86 | 92 | 94 | 99 |
| **Looking down** | 65 | 72 | 77 | 81 |
| **Looking up** | 42 | 53 | 63 | 83 |
| **Frontal gesture** | 75 | 78 | 82 | 92 |
| **Frontal laugh** | 70 | 72 | 80 | 90 |
| **Frontal smile** | 83 | 85 | 90 | 93 |

## 5 Conclusion

Fifty-three features representation have been extracted from twelve anchor points detected as in section 2. All features were obtained from angle and distance measurement. Each feature has its own discriminating power. From the discriminating power values, eye separation is the most significant features can be used to discriminate the matching process, followed by mean of mass distances. Span of eyes and several distance measurement such as distance nose tip to upper nose, nose tip to chin, and nose tip to right base nose are at the lower ranking. These features will then be tested to show which features will get the optimum performance. It means that not all features will be used for matching.

## REFERENCES

[1] Lee, M.W. & Ranganath, S. 2003. Pose-invariant face recognition using a 3D deformable model. *Pattern Recognition 36*, pp. 1853-1846.

[2] Huang, J., Heisele, B. & Blanz, V. 2003. Component-based Face Recognition with 3D Morphable Models. *4th Conf. on Audio and Video-based Person Authentication*, pp. 85 – 92.

[3] Gupta, S., Markey, M. K. & Bovik, A. C. 2007. Advances and challenges in 3D and 2D+3D human face recognition. *Pattern Recognition Research Horizons , F. Columbus (Editors) , Hauppauge, New York:* Nova Science Publishers.

[4] Gordon, G.G. 1992. Face Recognition based on depth and curvature features. *Computer Vision and Pattern Recognition (CVPR)*, pp. 108-110.

[5] Moreno, B., Sanchez, A., Velez, J. F. and Diaz, F. J. 2005. Face Recognition Using 3D Local Geometrical Features: PCA vs. SVM. *Image and Vision Computing*, pp. 339-352.

[6] Xu, C., Wang, Y., Tan, T. & Quan, L. 2004. Automatic 3D Face Recognition Combining Global Geometric Features with Local Shape Variation Information. *Sixth International Conference on Automated Face and Gesture Recognition, Seoul, Korea*, pp. 308-313.

[7] Chua , C-S., Han, F. & Ho, Y-K. 2000. 3D human face recognition using point signature. *Proceeding IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 233–238.

[8] Wang. Y., Chua., C.-S. & Ho., Y.-K. 2002. Facial feature detection and face recognition from 2D and 3D images. *Pattern Recognition Letters* **23**: 1191–1202.
[9] Wang, Y. & Chua, C-S. 2005. Face recognition from 2D and 3D images using 3D Gabor filters. *Image Vision Computing* **23**(11): 1018-1028.
[10] Bowyer, K., Chang, W. K. & Flynn, P. 2006. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding* (101): 1–15.
[11] Lu, X. & Jain, A. K. 2006. Deformation Modeling for Robust 3D Face Matching. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR2006), pp. 1377-1383.
[12] Moreno, B., Sanchez, A., Velez, J. F. and Diaz, F. J. 2003. Face Recognition using 3D surface-extracted descriptors. *In Proceedings of the Irish Machine Vision & Image Processing Conference 2003 (IMVIP'03).*

**Fatimah Khalid** completed PhD from University Kebangsaan Malaysia in 2008. Currently she is Lecturer at the Department of Multimedia, Faculty of Computer Science and Information Technology, University Putra Malaysia. Her research interest include image processing and computer vision.

**Lili Nurliyana** completed PhD from University Kebangsaan Malaysia in 2008. Currently she is Senior Lecturer at the department of Multimedia, Faculty of Computer Science and Information Technology, University Putra Malaysia. Her research interest includes video retrieval, computer game and animation.