# 3D Face Recognition using Statistical Multiple Features for the Local Depth Information


Yeunghak Lee[1], Kunwoo Park[1], Jaechang Shim[2], Taihong Yi[1]

*Yeungnam University[1], Andong University[2]*

*annaturu@yumail.ac.kr[1], kunwoosir@yumail.ac.kr[1]*

*jcshim@andong.ac.kr[2], taehong@yumail.ac.kr[1]*



## Abstract

*Depth information is one of the most important factors for the recognition of a digital face image. Range images are very useful, when comparing one face with another, because of implicating depth information. As the processing for the whole face produces a lot of calculations and data, face images can be represented in terms of a vector of feature descriptors for a local area. In this paper, depth areas of a 3 dimensional (3D) face image were extracted by the contour line from some depth value. These were resampled and stored in consecutive location in feature vector using multiple feature method. A comparison between two faces was made based on their distance in the feature space, using Euclidian distance. This paper reduced the number of index data in the database and used fewer feature vectors than other methods. Proposed algorithm can be highly recognized for using local depth information and less feature vectors on the face.*


## 1. Introduction

Today's computer environment is changing because of the development of intelligent interface and multimedia. To recognize the user automatically, people research various recognition methods using biometric information – fingerprint, face, iris, voice, vein, etc [1]. In a biometric identification system, because face recognition is a no-touch style, it is a very challenging research area, next to the fingerprint. It is influenced by lighting illuminance and encounters difficulties when the face is angled away from the camera. These factors cause low recognition. To solve these problems a computing company has developed a 3D face recognition system [2-3]. To obtain a 3D face, this method uses stereo matching, laser scanner, etc. Stereo matching extracts 3D information from the disparity of 2 pictures which is taken by 2 cameras. Even though it can extract 3D information from near and far away, it has many difficulties when used because of its low precision. 3D laser scanners extract more accurate depth information about the face, and because it uses a filter and a laser, it has the advantage of not being influence by the lighting illuminance when its angled away from the camera. Because it can measure distance, a 3D face image can be reduced by a scaling effect that is caused by the distance between the face and the camera [4].

Broadly speaking there are two ways to recognize the face feature based approach and the area based approach [4-7]. A feature based approach uses feature vectors which are extracted from within the image as a recognition parameter. An area based approach extracts a special area from the face and recognizes it using the relationship and minimum sum of squared difference. Face recognition research usually uses two dimensional images. Recently, the 3D system has become cheaper, smaller and faster to process than it used to be, so the use of 3D face image is now being more briskly researched [3][8-11]. Many researchers have used 3D face recognition using differential geometry tools for computing of curvature [8]. Hiromi et. al [9] treated the face recognition problem as a 3D shape recognition problem of rigid free-form surfaces. Each face in the input images and model database is represented as an Extended Gaussian Image (EGI), constructed by mapping principal curvatures and their directions. Gordon [10] presented a study of face recognition based on depth and curvature features. To find face specific descriptors, he used curvatures of the face. Comparison between the two faces was made based on their relationship in the feature space. Lee and Milios [12] extracted the convex regions of the face by segmenting the range of the images based on the sign of the mean and Gaussian curvature at each point. For each of these convex regions, the Extended Gaussian Image (EGI) is extracted and then used to match the facial features of the two face images.

In this paper, we will introduce a novel face recognition algorithm using multiple features for the area in contour line of face, which has depth information. The rest of the paper is structured as follows. Section 2 extracts the depth area for face. In Section 3, proposed multiple feature algorithms is presented. Experimental results for 3D images are shown in Section 4, and conclusions are given in Section 5.

### 2.1. Find the nose tip on the 3D face

During preprocessing, a given 3D image is divided into a face and background areas. Useless areas around the head and cloth parts include too much error data to process. In here, backgrounds have the lowest depth value in the image. Fig. 2 (a) shows the result of using Sobel Operator for the boundary line of the face, background and cloth, by equation (1),

| | | |
|---|---|---|
| d0<br>i-1, j-1 | d1<br>i-1, j | d2<br>i-1, j+1 |
| d4<br>i, j-1 | d5<br>i, j | d6<br>i, j+1 |
| d7<br>i+1, j-1 | d8<br>i+1, j-1 | d8<br>i+1, j+1 |

**Figure 1. 3 x 3 mask**

$$S_{xx} = d6 + 2*d7 + d8 - d0 - 2*d1 - d2$$
$$S_{yy} = d2 + 2*d5 + d8 - d0 - 2*d3 - d6$$

$$Th = \rho\sqrt{S_{xx}^2 + S_{yy}^2}$$

$$S_{img} = \begin{cases} 255, & Th > threshold \\ I_{img}, & else \end{cases} \quad (1)$$

where $\rho$ is 10 and the threshold is 255, ***I*** is input image and ***S*** is the result of the sobel image. Usually, because background and cloth is deeper than the face, the maximum value is extracted for 4 corners using a 5x5 window. If the pixel value is smaller than the maximum value, the pixel changes to 0, the result is Fig. 2 (b). This is the process to remove the noise caused by clothes. For the 3D image we removed the noise included in the clothe and background, and obtained a binary image using the average 3D image as a threshold value. The result is Fig. 2 (c).

$$Avg = \frac{1}{N}\sum S_{img}(i,j), \quad S_{img} \neq 255$$

$$B_{img} = \begin{cases} 0, & S_{img} > Avg \\ 255, & else \end{cases} \quad (2)$$

## 2. Extraction of contourlines

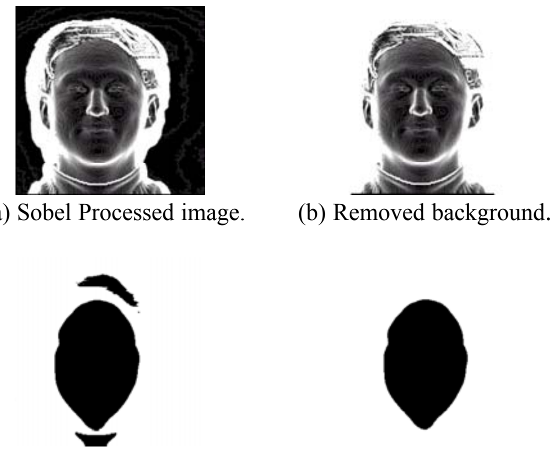
Because an extracted binary image includes noise information about hair, using an object labeling algorithm, the largest area is extracted in the binary image. The image in Fig. 2 (d) is the result of the removal of the background and clothe.

The face is a curved surface and can be divided into three parts: eyes, nose and mouth. Because the obtained image uses a 3D laser scanner, it has three dimensional depth information; the X, Y and Z coordinates, as in Fig. 3 (a). Even though the face is rotated and panned within some limited value, we find that the nose tip has its maximum value in the face. Fig. 3 (b) shows the nose tip is the maximum point from the side view. The nose is located in the center of the face and has maximum value. We can use the nose tip as a reference point, as it is easy to find in the face.

To find the maximum points, the average value is used as the threshold value, as de fined in (3). We can extract an area over the threshold value, as defined in (4)

$$avg = \frac{1}{M*N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} P(i,j), P(i,j) > 0 \quad (3)$$

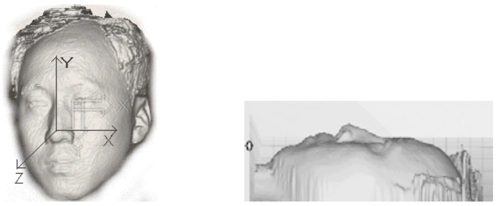
$$P(i,j) = \begin{cases} P(i,j), P(i,j) > avg \\ 0, Otherwise \end{cases} \quad (4)$$



(a) Sobel Processed image. (b) Removed background.

(c) Adapted threshold value. (d) Noise removed image.

**Figure 2. Preprocessing to remove background and noise.**




(a) 3 dimensional coordinate. (b) Side view image.

**Figure 3. Coordinates and side view of 3D face image.**

where M and N represent the horizontal and vertical size of the image. In the result area the new threshold value is calculated by using (3), and the new area is extracted by using (4). By repeating processes (3) and (4), we can find the nose tip point which has the highest value on the face. Because everyone's nose tip has a different shape, there are several pixels having a maximum value at the nose tip. Therefore by adapting the center of the weight around the maximum points, a nose tip is designated as the one maximum point, as shown in Fig. 4.


**Figure 4. The result of nose tip finding.**

## 2.2. Face normalization

In feature recognition of 3D faces, one has to take into consideration the obtained frontal posture. Face recognition systems suffer drastic losses in performance when the face is not correctly oriented. The normalization process proposed here is a sequential procedure that aims to put the face shapes in a standard spatial position. The processing sequence is panning, rotation and tilting. Firstly, for the panning, the face is panned by angle which is defined by the mean and distance of the depth value of local areas that are divided into right and left, as defined in (5), (6), (7) and (8). The window size used is 30x30, with the nose tip as the reference point. For example, the given binary images presented in Fig. 5 are before compensation (BC) and after compensation (AC) for panning.

$$X_L = \frac{1}{n}\sum_{i=s1}^{s1-30} x_i \quad \text{and} \quad X_R = \frac{1}{n}\sum_{i=s2}^{s1+30} x_i \tag{5}$$

$$M_L = \frac{1}{n}\sum_{i=s1}^{s1-30}\sum_{j=\max\_y}^{\max\_y-30} D_{img}, \quad \text{if } D_{img} > 0$$

$$M_R = \frac{1}{n}\sum_{i=s1}^{s1-30}\sum_{j=\max\_y}^{\max\_y+30} D_{img}, \quad \text{if } D_{img} > 0 \tag{6}$$

$$L = X_L - X_R \quad \text{and} \quad M = M_L - M_R \tag{7}$$

$$\theta = a\tan\left(\frac{M}{L}\right) \tag{8}$$

$X_L$ : centroid of left area
$X_R$ : centroid of right area
$M_L$ : mean value of left local area
$M_R$ : mean value of right local area
$L$ : difference of two centroid value
$M$ : difference of two mean value
$\theta$ : rotated angle
s1 : max_x-15
s2 : max_x+15

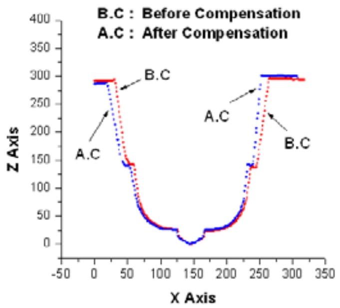



**Fig. 5 The processing of panning. BC is before compensation and AC is after compensation.**

Second, for the rotation, the face is rotated by angle which is defined by the modified centroid and moments used in several area [17]. In general, centroid and moments are given as the equivalent of their continuous counterparts. Given *B*, which is the binary image, a set of *n* pixels $p_i = (x_i, y_i)$ (i=1,…,*n*), the coordinates $(x_q, y_q)$ of the centroid q of *B* are calculated by

$$x_q^p = \frac{1}{n}\sum_{i=1}^{n} x_i^p \quad \text{and} \quad y_q^p = \frac{1}{n}\sum_{i=1}^{n} y_i^p \tag{9}$$

where *p* is contour values, p=15,16,17,18,19,20.

Moments allow for a unique characterization of a shape. Given *B*, a set of n pixels $p_i = (x_i, y_i)$ (i=1,…,n), and $q = (x_q, y_q)$ its centroid, the definition of the discrete (*k*, *l*)-order central moment $\Pi_{k,l}^p$ of the set B is given by

$$\Pi_{k,l}^p = \sum_{i=1}^{n}(x_i^p - x_q^p)^k (y_i^p - y_q^p)^l \tag{10}$$

A shape is uniquely represented by the set of all its (k,l)-ordered central moments (k, l $\in$ **N**). Clearly, the shape of the face is symmetrical with respect to the diagonal axis or nose bridge, $\Pi^p_{k,l} = \Pi^p_{l,k}$ for all $k \in$ **N** and $l \in$ **N**. Orientation is defined as an angle $\theta$ representing the angle made between the axis and the axis of least moment of inertia. The value of angle is obtained by the following formula (11).

$$\theta = \frac{1}{6}\sum_{p=15}^{20}\left(\frac{1}{2}a\tan\left(\frac{2\Pi^p_{1,1}}{\Pi^p_{2,0}-\Pi^p_{0,2}}\right)\right) \tag{11}$$

In Fig. 6, original image, rotated image and the binary images by the contour lines are presented.

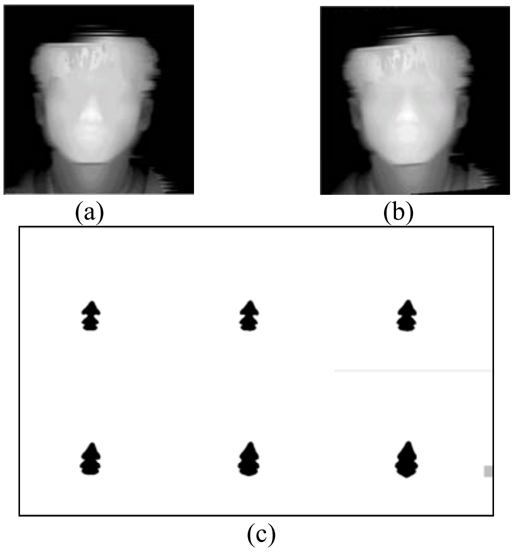



**Fig. 6 The processing of rotation (a) before rotation (b) after rotation (c) binary image by contour line p=15,16,17,18,19,20 from origin image.**

Third, for tilting, after finding the nose bridge and the nose base, the face is adjusted until the difference depth value of two points is became 10, as shown in Fig. 7. In Fig.7, the red profile line (B.C.) is before tilting and the blue profile line (A.C.) is after tilting.

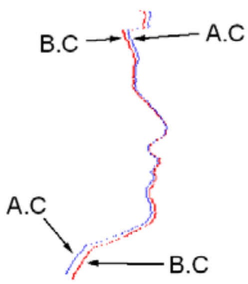



**Fig. 7 The processing of tilting. BC is before tilting and AC is after tilting.**

## 2.3. Extraction of the area in contour lines

Once the background has been removed and the 3D face normalized, we can take a binary image that is divided into inner and outer areas by the threshold contour line value (TCLV), which also stands for its depth value.

$$Th = \frac{1}{k}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} D_{img}(i,j) \quad \text{if } D_{img} > 0 \tag{12}$$

$$D_{img} = \begin{cases} D_{img} = D_{img}, & \text{if } D_{img} > Th \\ D_{img} = 0, & else \end{cases} \tag{13}$$

$$DB_{img} = \begin{cases} 0, & \text{if } D_{img} > TLCV \\ 255, & else \end{cases} \tag{14}$$

where $DB_{img}$ is binary image and $D$ is 3D image.

During this process, we adapt the iterative selection theory to find the contour area as defined in (12), (13) and (14). The initial $Th$ at the threshold is simply the mean of depth value, in (12), where k is the counter of pixels that is greater than 0. This threshold is then used to stop the processing of iterative. If the $Th$ is smaller than the TCLV, the process continues to (13) and the process is repeated. But if the $Th$ is greater than or equal to the TCLV, then the process stops. Finally the binary image is obtained by (14). Fig. 8 shows the results of the process. It also includes the nose and noise area. To remove the noise area, we used labeling process. Each area is labeled and the biggest area is extracted.

First, the labeling engine finds black pixels in the binary image. If the black pixel is connected with other pixels for 8 directions, the engine records the before labeled number. The engine also tries to find a connection between the pixel and the other black pixels. If it has no other connections, the engine stops searching in 8 directions. As the engine moves to the next pixel, it decides if the pixel is black or white. If the pixel is black, the label value is increased and it does this again before processing.

The results of various threshold values are shown in Fig. 9. Fig.9 (a) is the smallest area and Fig 9. (d) is the biggest area. These two areas are very difficult to compare for face shape because each person has a similar shape. In this study we used contour lines at a depth of 20 and a depth of 30, as shown in Fig. 9 (b) and Fig. 9 (c). The extracted area around the nose is reassigned to the center point of the image. Then finally binary image and 3D image are saved by the threshold values.

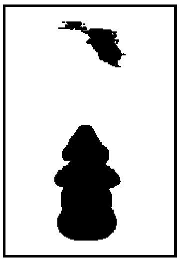
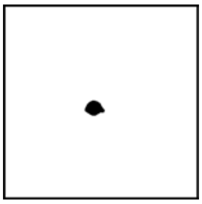


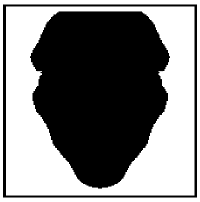
**Figure 5 Extracted areas using depth value 20.**



(a) Used depth value 10.

(b) Used depth value 20.

(c) Used depth value 30.



(d) Used depth value 40.

**Figure 6. Extracted binary images using depth values**

## 3. Multiple features for the area in contour line

There are two ways to use features to classify objects [13]. In statistical approaches many features are combined into a large feature vector. The same object can correspond to a wide variety of feature vectors just through the errors in measurements. However, these measurements will be clustered in some region of *N*-space. Hence a statistical recognizer will construct a feature vector from a data object and classify it based on how far. As a contrast, structural pattern recognition is that objects are constructed from smaller components using a set of rules. Structural pattern recognition is, in fact, a sophisticated variation on template matching, one that must match relationships between objects as well as the objects themselves. The problems involved in structural pattern recognition are two: locating the components, and finding a good representation for storing the relationships between the components. In general, there are many such features: the area, the perimeter, the circularity, the rectangularity, the size and the location of any holes etc. Multiple features consist of a collection of values, and are themselves vectors. In general, the projection executes the role of converting N dimensional coordinate systems into smaller dimensions than N. For example, to reample a 12 x 12 shape to size of 3 x 3, simply group the pixels into three groups of four in each direction, as shown in figure 7.

In this paper, resampled image divided by 5x5 was used as feature vectors. Used Features were the average and variance of depth value for each group. It is possible to suitably adjust the number of blocks for precision, so the system can reduce index data and obtain good results. In this study, we used 5x5 groups for images, for a total of 25 areas. We used feature vectors, as given by

$$Avg(k) = \frac{1}{M_k * N_k} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (P - R_{img}(i,j)) \tag{15}$$

$$Var(k) = \frac{1}{M_k * N_k} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (Avg(k) - (P - R_{img}(i,j)))^2 \tag{16}$$

$$(\text{if } B_{img}(i,j) = 0)$$

where $\boldsymbol{P}$ is the value of nose peak, $\boldsymbol{R}$ is range image, k=1,…25 and $\boldsymbol{B}$ is the binary image of contour line.

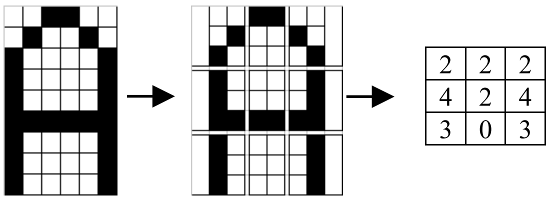



**Figure 7. Resampling of a shape to provide a multiple feature.**

## 4. Experimental results

In this study, we used a 3D laser scanner made by a 4D Culture to obtain a 3D face image. First, a laser line beam was used to strip the face for 3 seconds, thereby obtaining a laser profile image that is, 180 pieces. Obtained image size is extracted by using the extraction algorithm of line of center, is 640x480. Next, calibration is done in order to process the height value, resampling and interpolation. Finally, a 3D face image using this paper is extracted, at 320x320.

From this 3D face image, finding the nose tip point, using contour line threshold values at 40 (for which the fiducial point is nose tip), we extract images included around the nose area. After the indexing system is constructed by using multiple feature vectors which are adapted from the same size of block for each image, recognition results are compared and queried with the database system. The similarity of indexing system used L2-norm as given by (17), (18) and (19) for contour line threshold values.

$$D_{Avg}(i)=\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{(I_{Avg}(j)-Q_{Avg}(i,j))^2} \quad (17)$$

$$D_{Var}(i)=\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{(I_{Var}(j)-Q_{Var}(i,j))^2} \quad (18)$$

$$S=\min_{k=1 \text{ to } M}\left\{\sqrt{D_{Avg}^2(k)+D_{VAr}^2(k)}\right\} \quad (19)$$

***M*** is database size and ***N*** is the size of block. ***I*** is the input image and ***Q*** is the database images. ***S*** is the result of similarity between input and database image.

A database is used to compare the different strategies and is composed of 70 images (two images of 35 people). Of the two pictures available, after the first were taken, the second photos were taken at time intervals of 30 minutes. The results then ranked threshold values arranging them, in ascending order in accordance with their measured similarity with respect to the current input. We showed the result of recognition for the combination of each area (Table 1 and Fig. 8). Specifically, for the area 40, the matching face was included among the best 5 and 10 ranked candidates in 94.3% and 100% of the cases, respectively (Table 1. and Fig. 8). And Fig. 9 showed for the result of direct matching method about each area. This means that in the contour line 20 and 30 were difficult to distinguish query image from model images, because the small area of face is very similar. And direct matching is also more difficult to recognize than multiple feature method. In Takacs and Wechsler's [14] experiment, because they used whole face and Hausdorff distance algorithm to compare similarity, this system needed very complex calculation to recognize. But in this study, because we used local depth areas on the face and multiple feature vectors as parameters, parameters and calculation were less than others. The recognition result was a similar recognition rate with Takacs and Wechsler. Regarding the large database, we expect excellent recognition results by using the statistical feature of multiple feature vectors.

**Table 1. The rate of recognition for each contour line.**

| | Best 5 | Best 10 | Best 15 |
|---|---|---|---|
| Contour line 20 | 80.0% | 86.0% | 89.0% |
| Contour line 30 | 85.7% | 94.0% | 97.0% |
| Contour line 40 | 94.3% | 100% | 100% |

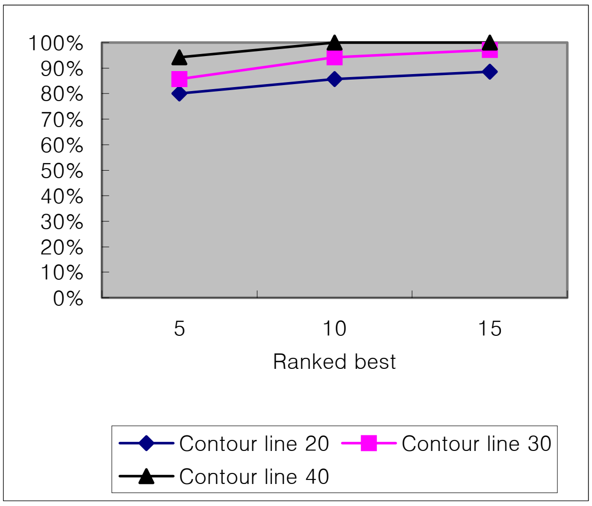



**Figure 8. Recognition rate for multiple feature method.**

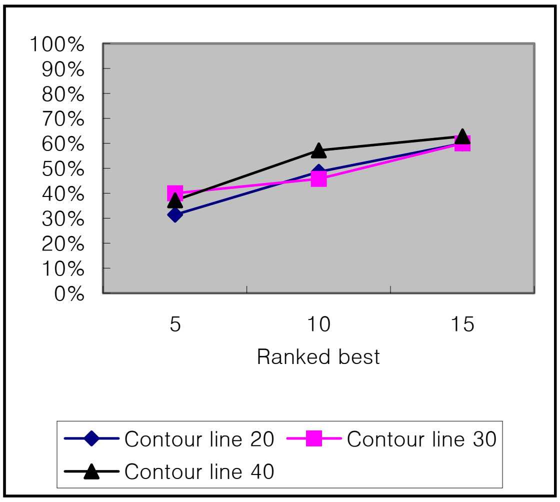



**Figure 9. Recognition rate for direct matching method.**

## 5. Conclusion

We have introduced, in this paper, a new practical implementation of a person verification system using the shape of 3-dimensional (3D) face images based on multiple feature method. The underlying motivations for our approach originate from the observation that the outline of the face has different shape and has different depth values according to the some depth value from the nose tip.

We found the exact nose tip point by using a very simple algorithm and made contour areas using iterative selection theory for the depth of the face. The local area around the nose was extracted by using contour lines threshold. The statistical feature of multiple feature vectors represented was too robust for the local area of the face. Experimental results on a group of face images (70 images) demonstrated that our approach produces excellent recognition results. The matching face was included among the best 5 and 10 ranked candidates in 94.3% and 100% of the cases, respectively.

These results prove that the process of face recognition may use less parameters and calculations than earlier suggested. We argue, that even though a large database is constructed, if the statistical features of projection vectors are used, indexing time and complexity will be reduced because of fewer parameters.

## 6. References


[1] L. C. Jain, U. Halici, I. Hayashi, S. B. Lee, *Intelligent biometric techniques in fingerprint and face recognition,* CRC Press, 1999.

[2] 4D Culture, "http://www.4dculture.com".

[3] Cyberware, "http://www.cyberware.com".

[4] P. L. Hallinan, G. G. Gordon, A. L. Yuille, P. Giblin, D. Mumford, *Two and three dimensional pattern of the face*, A K Peters. Ltd. 1999.

[5] M. Grob, *Visual computing*, Springer_Verlag, 1994.

[6] A. Nikolaidis, I. Pitas, "Facial feature extraction and pose determination," *Pattern Recognition*, 33, pp. 1783-1791, 2000.

[7] B. Moghaddam, T. Jebara, A. Pentland "Bayesian face recognition," *Pattern Recognition*, 33, pp. 1771-1782, 2000.

[8] C. S. Chua, F. Han, Y. K. Ho, "3D Human Face Recognition Using Point Signature," *4th ICAFGR*, 2000.

[9] H. T. Tanaka, M. Ikeda and H. Chiaki, "Curvature-based face surface recognition using spherial correlation," *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 372-377, 1998.

[10] G. G. Gordon, "Face Recognition based on depth and curvature feature," *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 808-810, 1992.

[11] R. Chellapa, C. L. Wilson, and S. Sirohey "Human and machine recognition of faces: A survey," *Proceeding of the IEEE*, 83(5), 705-740, 1995.

[12] J. C. Lee and E. Milios. "Matching range image of human faces," *Third International Conference on Computer Vision*, pp. 722-726, 1990.

[13] J. R. Parker, *Algorithm for image processing and computer vision*, John Wiley & sons, Inc., 1997.

[14] Barnabás Takács, Harry Wechsler, "Face Recognition Using Binary Image Metrics", Proc. of the 3rd Int. Conf. on Automatic Face and Recognition, Nara, Japan, pp. 294-299, 1998.