3D Face Recognition Using Orientation Maps

B.H. Shekar[1], N. Harivinod[1], and M. Sharmila Kumari[2]

[1] Department of Computer Science, Mangalore University, Karnataka, India
{bhshekar,harivinodn}@gmail.com
[2] Department of Computer Science and Engineering, P.A. College of Engineering, Mangalore, Karnataka, India
sharmilabp@gmail.com

Abstract. In this work we present a new 3D face recognition method based on orientation maps. The proposed model consists of a method for extracting distinctive features from range images of face that can be used to perform reliable matching between different poses of a face. For a 3D face scan, range image is computed and the potential interest points are identified by searching at all scales. Based on the stability of the interest point, significant points are extracted. For each significant point, we compute the significant point descriptor which consists of vector made of values from the convolved orientation maps located on concentric circles centred on the significant point, and where the amount of Gaussian smoothing is proportional to the radii of the circles. Experiments have been conducted on the standard 3D face image database. Experiments show that the newly proposed method provides higher recognition rate compared to other existing contemporary models developed for 3D face recognition.

Keywords: Range Image, Local descriptor, Orientation map, 3D face recognition.

1 Introduction

Automatic face recognition systems have wide range of applications in access control, surveillance, personal identification and pervasive computing. Although 2D intensity based face recognition is easy to acquire and process, it is vulnerable to the change of pose and illumination. Hence much of the research in the recent days focused on 3D model based face recognition. 3D model based representation of faces provides invariance to illumination and pose changes.

The 3D model based face recognition techniques are categorized based on the global information or local information of range images. In [10], Heseltine et al. applied the principle component analysis directly to the range images and used the Euclidean distance to measure similarities. In [11], Hesher et al. made statistical analysis using principle component analysis and independent component analysis, and then impose probability models on the coefficients. Achermann et al. [2] utilized the eigen face and Hidden Markov Model for recognition on range images. In [16], Moreno et al. analyze range image based face recognition on

A. Abraham et al. (Eds.): ACC 2011, Part III, CCIS 192, pp. 396–405, 2011.

three classes: full, upper-half and left-side facial depth map. Two-dimensional FLD based face recognition was proposed by Guru and Vikram [9].

On the other hand, we have seen the local feature based point descriptors such as *Scale Invariant Feature Transform(SIFT)* and *Speeded Up Robust Features(SURF)*, that are widely used in the area of face recognition. It is shown that these features are invariant to affine transformations and illumination changes. In [15], Lowe introduced SIFT to perform matching between different views of an object. This 128 dimensional descriptor is based on the local image gradient, transformed according to the orientation of the significant point to provide orientation invariance. Mian et al. [17] used SIFT descriptors for face recognition under illumination and expression variations by combining the 2D and 3D local feature. In [8][14], Guo et al. and Lo et al. used 2.5D SIFT Descriptor for facial feature extraction in range images. In [7], Geng and Jiangs use variant of SIFT called volume-SIFT and partial-descriptor-SIFT for face recognition on 2D faces. In [13], the SIFT descriptors are computed at specific points on a regular grid of the face image and robustness to illumination variations is achieved. Bay et al. in [3] presented SURF as a detector and descriptor. Kim and Dahyot [12] uses SURF for face components detection using support vector machines. Yunqi et al. used SURF for 2D and 3D face recognition [20][21]. In [1], An et al. used SURF for face detection and recognition with SURF for human-robot Interaction. Tola in [18] proposed DAISY, an efficient dense descriptor applied to wide-baseline stereo. This descriptor is inspired by SIFT and is computationally efficient. Velardo et al. in [19] applied this work to face recognition on 2D images.

Most of the works discussed above use the 3D data in the form of point clouds. Here, the 3D sensors, used for face capture, produces 2.5D information [6]. This data can be easily projected to a 2D image plane and is called *depth image* or *range image*. A range image is a 2D image in which each pixel represents the distance from a point of the face surface to a plane. It is observed that the range image construction normally preceded with a pose normalization phase in order to transform faces to a frontal configuration. We use *Iterative Closest Point (ICP)* algorithm for this purpose. Using interpolation methods, irregularly sampled 3D points are converted into regular *(x, y)* grid. Instead of comparing the whole image, we choose significant point comparison to find the match between two range images. The significant points are detected using Hessian based detector followed by the significant point descriptor computation. The significant point descriptor for all significant points are computed. The test face is matched with all images in the dataset. Two range images are said to be matched if they have maximum number of matching significant points.

The rest of the paper is organized as follows: The methodology of proposed 3D face recognition model is given in section 2. It describes the preprocessing, significant point extraction, significant point descriptor computation and matching. Database description and experimental results are provided in section 3. Conclusion is presented in section 4.

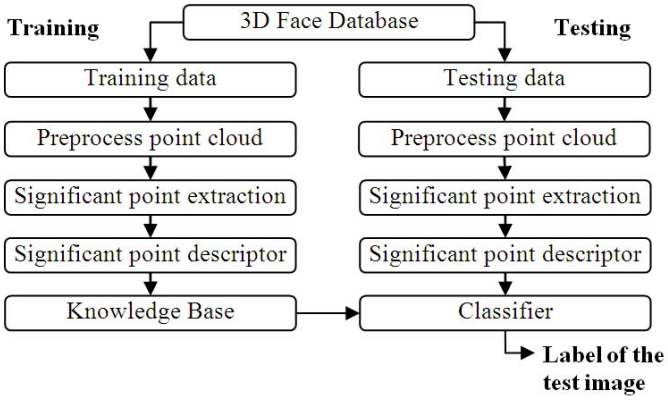

Fig. 1. Block diagram of the proposed 3D face recognition system

2 Methodology

The proposed method mainly consists of following steps: significant point extraction from the preprocessed range image, significant point descriptor computation and matching. The block diagram of the proposed face recognition system is given in Figure 1.

2.1 Preprocessing

The 3D face point clouds are obtained from a 3D face database [6]. Since the face scans of a person differ with pose, they need to be aligned. We use ICP [4] to do the registration. It performs the registration operation automatically without any human intervention. In our experiments all the scans of persons are aligned with the first frontal scan of that person.

The face scans usually do not contain corresponding data with respect to the each grid in the range image. So the scattered data is linearly interpolated. We interpolated the data using a Delaunay triangulation. Then the nose tip is identified. In our implementation, the point nearer to the scanner is identified as nose tip which posses highest depth value. Since all the scans are aligned to the frontal scan, the value with the highest depth value corresponds to the nose tip. Keeping nose tip as the center, the range image is cropped elliptically. The range images before preprocessing are shown in the left side of Figure 2 and on the right side are the range images obtained due to preprocessing.

2.2 Significant Point Extraction

The significant points are extracted using the SURF detector [3] which is based on the determinant of the Hessian matrix. We have used integral images to reduce the computational burden of the significant point extraction. Integral image helps in fast computation of sum of the rectangular region in an image.

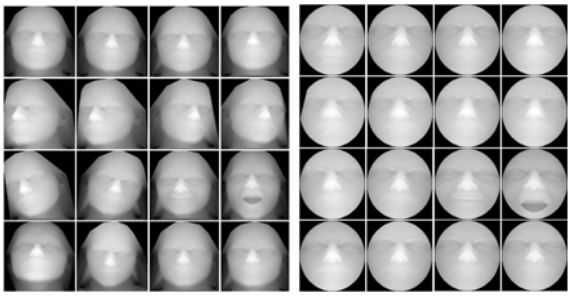

Fig. 2. All range images of a person before(left) and after(right) preprocessing of the point cloud

Due to this box type convolution filters are computed quickly. For an image I, the integral image I_Σ is defined as follows.

$$I_\Sigma(x,y) = \sum_{i=1}^{i\leq x}\sum_{j=1}^{j\leq y} I(i,j) \tag{1}$$

where x and y represent the row and column number of a pixel in image I. $I_\Sigma(x,y)$ is the sum of all $I(x,y)$ terms to the left and above the pixel (x,y). Using integral images, the sum of the rectangular region in the image I can be calculated using three arithmetic operations [3].

Hessian based SURF detector is chosen because it is more stable and repeatable. It uses the determinant of the Hessian matrix. For a point $p = (x,y)$ in image I, the Hessian matrix $H(p,\sigma)$ is the matrix of partial derivatives of the image I, in the following form,

$$H(p,\sigma) = \begin{bmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{xy}(p,\sigma) & L_{yy}(p,\sigma) \end{bmatrix} \tag{2}$$

where $L_{xx}(p,\sigma)$ is the convolution of the Gaussian second order derivative with the image I in point p, and similarly for $L_{xy}(p,\sigma)$ and $L_{yy}(p,\sigma)$. Gaussians are widely used for scale-space analysis [3]. Its discrete formations are used in actual implementation. The first two diagrams in Figure 3 show discrete Gaussian second order derivative. The Hessian matrix is computed using the box filters which approximate second order Gaussian derivatives. The last two diagrams in Figure 3 illustrate the same. The 9x9 box filters in Figure 3 are approximations of a Gaussian with $\Sigma = 1.2$ and represent the lowest scale for computing the blob response maps. Suppose D_{xx}, D_{yy} and D_{xy} are the approximations to I_{xx}, I_{yy} and I_{xy}, the determinant of H_{approx} is computed as,

$$det(H_{approx}) = D_{xy}D_{xy} - (wD_{xy})^2 \tag{3}$$

where w is the weight of the filter response used to balance the expression for the Hessian's determinant. The approximated determinant of the Hessian represents the blob response in the image at location p.

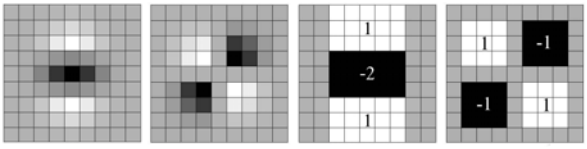

Fig. 3. The first two images represent the discretised Gaussian second order partial derivative in y (L_{yy}) and xy direction (L_{xy}) respectively. The last two images represent the approximation for the second order Gaussian partial derivative in y (D_{yy}) and xy direction (D_{xy}). The gray regions assumed to be zero. (Image Courtesy [3]).

For a given image, the potential interest points are identified by searching at all scales and based on the stability of the interest point, significant points are extracted. Lowe [15] implemented the scale spaces using the image pyramid. Each higher layer in the pyramid contains subsampled, smoothed image with Gaussian kernels. Here the image size is reduced due to subsampling. But Bay et al. [3], constructed the scale space by increasing the filter size keeping the image size fixed. They used the box filters and integral images. Box filters of any size can be applied on the original image at same speed. We employ the technique proposed in [3].

The scale space construction followed by the non-maximum suppression in the neighborhood is performed. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space with the method by Brown et al. [5]. This represents the significant point in the range image.

2.3 Significant Point Descriptor Computation

Significant point descriptor computation is similar to descriptor which was introduced by Tola et al. in [18] for matching wide-baseline image pairs. In our work, for each significant point, we compute the descriptor which consists of vector made of values from the convolved orientation maps located on concentric circles centered on the significant point, and where the amount of Gaussian smoothing is proportional to the radii of the circles.

Significant point descriptor building process is divided into three stages: computing orientation maps, convolving orientation maps with Gaussian kernels, and concatenating significant point descriptor by reading the values from convolved response maps.

For each significant point in the image I, we first compute eight orientation maps, one for each quantized direction. Orientation map $G_i(u, v)$ for the significant point (u, v), equals the image gradient at (u, v) in the direction i, if it is greater than zero else it is equal to zero. $G_i = (\delta I/\delta i)^+, 1 \leq i \leq H$, where H is the total orientation maps and $(.)^+$ is the operator such that $(a)^+ = max\,(a, 0)$.

Each orientation map is then convolved several times with Gaussian kernels of different Σ values to obtain convolved orientation maps for different sized

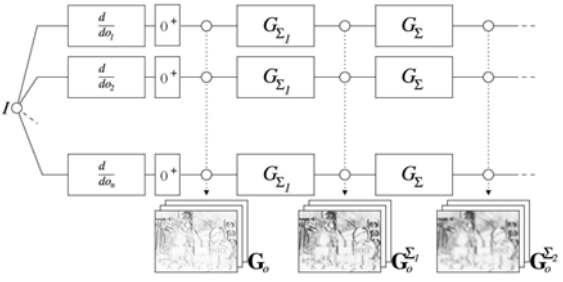

Fig. 4. Original image (I), Orientation maps (G_o) and Convolved orientation maps ($G_o^{\Sigma_i}$) . (Courtesy [18]).

regions as $G_i^{\Sigma} = G_{\Sigma} * (\delta I/\delta i)^+$with G_{Σ} a Gaussian Kernel. Different Σs are used to control the size of the region. This can be done efficiently by computing these convolutions recursively. Figure 4 shows these computations.

$$G_i^{\Sigma_2} = G_{\Sigma_2} * (\delta I/\delta i)^+ = G_{\Sigma} * G_{\Sigma_1} * (\delta I/\delta i)^+ = G_{\Sigma} * G_i^{\Sigma_1} \tag{4}$$

where $\Sigma = \sqrt{\Sigma_2^2 - \Sigma_1^2}$ and $\Sigma_2 > \Sigma_1$.

The descriptor for every significant point is computed by picking the values from the convolved response maps. As depicted in Figure 5, at significant point location, say (u, v), descriptor consists of vectors sampled in the neighborhood around it. These samples located on concentric circles and their amount of Gaussian smoothing is proportional to the radius of these circles. Let $h_{\Sigma}(u, v)$ be the vector made up of the values at location (u, v) in the convolved response maps.

$$h_{\Sigma}(u, v) = \left[G_1^{\Sigma}(u, v), G_2^{\Sigma}(u, v), ..., G_8^{\Sigma}(u, v)\right]^T \tag{5}$$

where $G_1^{\Sigma}, G_2^{\Sigma}, ..., G_8^{\Sigma}$ denote the Σ convolved response maps. Before concatenating these vectors to a descriptor, we normalize them to unit vector, and denote the normalized vectors by $\tilde{h_{\Sigma}}(u, v)$. The full descriptor $D(u, v)$ for the significant point location (u, v) can be defined as a concatenation of vectors and can be written as:

$$\begin{aligned} D(u, v) = [&\tilde{h}_{\Sigma_1}^T(u, v), \\ &\tilde{h}_{\Sigma_1}^T(I_1(u, v, R_1)), ..., \tilde{h}_{\Sigma_1}^T(I_N(u, v, R_1)), \\ &\tilde{h}_{\Sigma_2}^T(I_1(u, v, R_2)), ..., \tilde{h}_{\Sigma_2}^T(I_N(u, v, R_2)), \\ &... \\ &\tilde{h}_{\Sigma_Q}^T(I_1(u, v, R_Q)), ..., \tilde{h}_{\Sigma_Q}^T(I_N(u, v, R_Q))]^T \end{aligned} \tag{6}$$

where $I_j(u, v, R_1)$ is the location with distance R from (u, v) in the direction given by j when the directions are quantized into N values. Figure 5 shows the sample locations when $N = 8$, and significant point descriptor is made up of values extracted from 25 locations and 8 response maps. Therefore, descriptor length is 200 (i.e. 8×25).

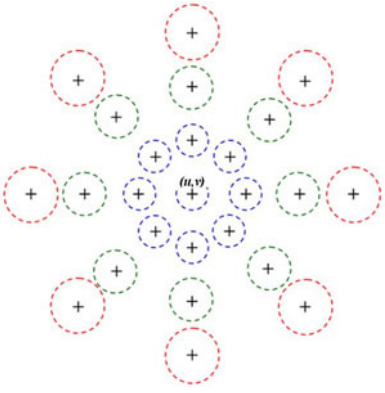

Fig. 5. Shape of the sampling locations of the descriptor. The '+' sign indicates the sampling locations. The radius of the dashed circle represents the size of the Gaussian kernel.

2.4 Similarity between Two Face Images

Once we have represented all face images as a set of interest points and their corresponding descriptions, the next step to be carried out is to find the similarity measure between two face images, in order to decide whether such images correspond to the same person or not. The comparison of two images is done by comparing the significant points. For each point of the first image, the best and second best matching points of the second image must be found. If the first match is much better than the second one, the points are said to be alike. Equation 7 shows how to apply such condition, where points B and C in range image I_2 are the best and second best matches, respectively, for point A in range image I_1.

$$\frac{\left|D^A_{I_1} - D^B_{I_2}\right|}{\left|D^A_{I_1} - D^C_{I_2}\right|} < threshold \tag{7}$$

3 Experimental Results

3.1 Database

We conducted experiments using FRAV3D face database [6]. It contains the 3D point clouds of 106 persons with 16 scans per each person. This includes facial scans with frontal (1,2,3,4), 25 right turn in Y direction (5,6), 5 left turn in Y direction (7,8), severe right turn in Z direction (9), small right turn in Z direction (10), smiling gesture (11), open mouth gesture (12), looking up turn in X direction (13), looking down turn in X direction (14), frontal images with uncontrolled illumination (15,16). The 2D image of all 16 scan of a subject is shown in left image of the Figure 2.

Table 1. Comparison of recognition accuracy of the proposed model with subspace analysis based techniques

Test configuration	Total persons	Training samples	Testing samples	Samples tested	Recognition Rate		
					$(2d)^2$ FLD	PCA	Proposed Model
T1	90	1,2,3	4	90	96.67	95.56	98.89
T2	90	1,2,3,4	11	90	85.56	88.89	84.44
T3	90	1,2,3,4	12	90	52.22	46.67	81.11
T4	90	1,2,3,4	15.16	180	94.44	97.22	100.00
T5	90	1,2,3,4	7.8	180	94.44	87.22	90.00

Table 2. Comparison of recognition accuracy of the proposed model with local descriptor based techniques

Test configuration	Recognition Rate		
	SIFT	SURF	Proposed Model
T1	95.56	98.89	100.0
T2	80.00	84.44	90.00
T3	78.89	78.89	90.00
T4	96.67	100.0	100.0
T5	81.11	90.56	93.89

3.2 Results

We compared recognition accuracy of the proposed model with subspace analysis based techniques and local descriptor based techniques. Experiments are conducted with different test configurations.

In the subspace analysis based techniques, we used $(2D)^2FLD$ and conventional PCA directly on range image. Table 1 shows the results of various combinations of training and testing samples using these techniques. We used range images of frontal face scans in training for all test configurations. In test configuration T1, high recognition rate is observed because training and testing data contains only the frontal scan. In T2 and T3, test input contains a gesture. In T4, it is observed that illumination variation does not affect the 3D face recognition. Compared to T1, T5 results have less recognition rate due to self occlusion. On an average, it shall be observed from Table 1 that the proposed model outperform the subspace analysis based techniques.

In local descriptor based techniques we choose SIFT and SURF descriptor based algorithms proposed for 3D face recognition. Here the average number of significant points per face range image is fixed to 24. The experiments were conducted using the same test configuration as given in Table 1. Table 2 shows these results. It is observed that proposed model gives better accuracy with fewer number of significant points when compared to SIFT and SURF.

Experiments are also conducted using the leave-one-out strategy taking all 16 face scans of subjects. The results are given in Table 3. The recognition rate

Table 3. Comparison of proposed model using leave one out strategy

Total subjects	Samples tested	Recognition rate				
		$(2d)^2$ FLD	PCA	SIFT	SURF	Proposed model
10	160	91.25	93.12	91.25	96.25	98.75
20	320	87.81	91.56	92.50	96.25	95.31
30	480	86.04	90.00	91.88	95.83	94.79
40	640	84.84	87.81	90.00	94.38	94.69
90(All)	1440	81.60	86.18	87.22	91.84	92.85

decreases with the increase of persons used for experiments. It is observed that the test face with open mouth mismatches most of the times. Here also, the proposed method outperforms the other methods.

4 Conclusion

We have developed an algorithm for 3D face recognition based on the significant points described by orientation maps. Experimental results show that, the proposed model out performs the conventional holistic face recognition techniques and other local descriptor based models.

Acknowledgement

The authors would like to thank the support provided by DST-RFBR, Govt. of India vide Ref. No. INT/RFBR/P-48 dated 19.06.2009.

References

1. An, S., Ma, X., Song, R., Li, Y.: Face detection and recognition with SURF for human-robot interaction. In: IEEE International Conference on Automation and Logistics, pp. 1946–1951 (2009)
2. Achermann, B., Jiang, X., Bunke, H.: Face recognition using range images. In: International Conference on Virtual Systems and Multi Media, pp. 129–136 (1997)
3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding 110(3), 346–359 (2008)
4. Besl, P.J., McKay, H.D.: A method for registration of 3-D shapes. IEEE Transactions on PAMI 14(2), 239–256 (1992)
5. Brown, M., Lowe, D.: Invariant features from interest point groups. In: British Machine Vision Conference (2002)
6. FRAV3D Face database, `http://www.frav.es/databases/FRAV3d/`
7. Geng, C., Jiang, X.: Face recognition using sift features. In: 16th IEEE International Conference on Image Processing, pp. 3313–3316 (2009)
8. Guo, H., Zhang, K., Jia, Q.: 2.5D SIFT Descriptor for Facial Feature Extraction. In: 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 1235–1250 (2010)

9. Guru, D.S., Vikram, T.N.: 2D Pairwise FLD: A robust methodology for face recognition. In: IEEE Workshop on Automatic Identification Advanced Technologies, pp. 99–102 (2007)
10. Heseltine, T., Pears, N., Austin, J.: Three-dimensional face recognition: An eigensurface approach. In: ICIP, pp. 1421–1424 (2004)
11. Hesher, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range imaging. In: International Symposium on Signal Processing and Its Applications, vol. 2, pp. 201–204 (2003)
12. Kim, D., Dahyot, R.: Face components detection using SURF descriptor and SVMs. In: Machine Vision and Image Processing Conference, pp. 51–56 (2008)
13. Krizaj, J., Struc, V., Pavesic, N.: Adaptation of SIFT features for face recognition under varying illumination. In: MIPRO, Proceedings of the 33rd International Convention, pp. 691–694 (2010)
14. Lo, T.R., Siebert, J.P.: Local feature extraction and matching on range images: 2.5D SIFT. Computer Vision and Image Understanding 113(12), 1235–1250 (2009)
15. Lowe, D.G.: Distinctive Image Features from Scale Invariant Keypoints. International Journal of Computer Vision 20(2), 91–110 (2004)
16. Moreno, A.B., Sanchez, A., Velez, J.F.: Voxel-based 3d face representations for recognition. In: 12th International Workshop on Systems, Signals and Image Processing, pp. 285–289 (2005)
17. Mian, A., Bennamoun, M., Owens, R.: Face Recognition Using 2D and 3D Multimodal Local Features. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Nefian, A., Meenakshisundaram, G., Pascucci, V., Zara, J., Molineros, J., Theisel, H., Malzbender, T. (eds.) ISVC 2006. LNCS, vol. 4291, pp. 860–870. Springer, Heidelberg (2006)
18. Tola, E., Lepetit, V., Fua, P.: DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 815–830 (2010)
19. Velardo, C., Dugelay, J.: Face recognition with DAISY descriptors. In: 12th ACM Workshop on Multimedia and Security, pp. 95–100 (2010)
20. Yunqi, L., Xutuan, J., Zhenxiang, S., Dongjie, C., Qingmin, L.: Face Recognition Method Based on SURF Feature. In: International Symposium on Computer Network and Multimedia Technology, pp. 1–4 (2009)
21. Yunqi, L., Haibin, L., Xutuan, J.: 3D face recognition by SURF operator based on depth image. In: 3rd IEEE International Conference on Computer Science and Information Technology, vol. 9, pp. 240–244 (2010)