

- [4] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, Washington, DC, May 2002, pp. 734–742.
- [5] L. Wang, H. Ning, W. Hu, and T. Tan, "Gait recognition based on procrustes shape analysis," in *Proc. Int. Conf. Image Processing*, 2002, pp. 433–436.
- [6] L. Wang, H. Ning, T. Tan, and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 2, pp. 149–158, Feb. 2004.
- [7] D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic extraction and description of human gait models for recognition purposes," in *Comput. Vis. Image Understand.*, Apr. 2003, vol. 90, pp. 1–41.
- [8] P. J. Phillips, S. Sarkar, I. R. Vega, P. Grother, and K. W. Bowyer, "The gait identification challenge problem: Data sets and baseline algorithm," in *Proc. Int. Conf. Pattern Recognition*, Quebec City, QC, Canada, Aug. 2002, vol. 1, pp. 385–388.
- [9] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The human ID gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.
- [10] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa, "Identification of humans using gait," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1163–1173, Sep. 2004.
- [11] S. D. Mowbray and M. S. Nixon, "Automatic gait recognition via Fourier descriptors of deformable objects," in *Proc. Audio Visual Biometric Person Authentication*, 2003, pp. 566–573.
- [12] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, "An angular transform of gait sequences for gait assisted recognition," in *Proc. IEEE Int. Conf. Image Processing*, Singapore, Oct. 2004, pp. 857–860.
- [13] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, "Gait recognition using linear time normalization," *Pattern Recognit.*, vol. 39, pp. 969–979, 2006.
- [14] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, "Gait recognition using dynamic time warping," in *Proc. IEEE 6th Workshop on Multimedia Signal Processing*, Sep. 29–Oct. 1, 2004, pp. 263–266.
- [15] A. Kale, N. Cuntoor, A. N. Rajagopalan, B. Yegnanarayana, and R. Chellappa, "Gait analysis for human identification," presented at the 3rd Int. Conf. Audio and Video Based Person Authentication, Jun. 2003.
- [16] Y. Liu, R. Collins, and Y. Tsing, "Gait sequence analysis using frieze patterns," in *Proc. Eur. Conf. Computer Vision*, 2002, pp. 657–671.
- [17] P. Salembier, A. Oliveras, and L. Garrido, "Anti-extensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998.
- [18] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with HSV color information," in *Proc. IEEE Intelligent Transportation Systems*, 2001, pp. 334–339.
- [19] D. D. Hoffman and M. Singh, "Saliency of visual parts," *Cognition*, vol. 63, pp. 29–78, 1997.
- [20] K. Moustakas, D. Tzovaras, and M. G. Strintzis, "SQ-Map: Efficient layered collision detection and haptic rendering," *IEEE Trans. Visual Comput. Graphics*, vol. 13, no. 1, pp. 80–93, Jan./Feb. 2007.
- [21] D. Simitopoulos, D. E. Koutsonanos, and M. G. Strintzis, "Robust image watermarking based on generalized Radon transformations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 732–745, Aug. 2003.
- [22] P. Daras, D. Zarpalas, D. Tzovaras, and M. G. Strintzis, "Efficient 3-D model search and retrieval using generalized 3-D Radon transforms," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 101–114, Feb. 2006.
- [23] N. V. Boulgouris and Z. X. Chi, "Gait recognition using Radon transform and linear discriminant analysis," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 731–740, 2007.
- [24] J. Shutler and M. S. Nixon, "Zernike velocity moments for sequence-based description of moving features," *Image Vis. Comput.*, vol. 24, no. 4, pp. 343–356, 2006.
- [25] P. T. Yap, R. Paramesran, and S. H. Ong, "Image analysis by Krawtchouk moments," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1367–1377, Nov. 2003.
- [26] A. Mademlis, A. Axenopoulos, P. Daras, D. Tzovaras, and M. G. Strintzis, "3D content-based search based on 3D Krawtchouk moments," in *Proc. 3DPVT 2006*. Chapel Hill: Univ. North Carolina, 2006.
- [27] J. Little and J. Boyd, "Recognizing people by their gait: The shape of motion," *Videre: J. Comput. Vis. Res.*, vol. 1, no. 2, pp. 1–32, 1998.
- [28] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
- [29] I. G. Damousis, A. G. Bakirtzis, and P. S. Dokopoulos, "Network-Constrained economic dispatch using real-coded genetic algorithm," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 198–205, Feb. 2003.
- [30] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1–3, pp. 7–42, Apr./Jun. 2002.
- [31] L. Zongyi and S. Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 6, pp. 863–876, Jun. 2006.
- [32] L. Lee, G. Dalley, and K. Tieu, "Learning pedestrian models for silhouette refinement," in *Proc. Int. Conf. Comput. Vis.*, 2003, pp. 663–670.
- [33] D. Tolliver and R. Collins, "Gait shape estimation for identification," in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 2003, pp. 734–742.

### 3-D Face Recognition Using Local Appearance-Based Models

Hazým Kemal Ekenel, Hua Gao, and Rainer Stiefelhagen

**Abstract**—In this paper, we present a local appearance-based approach for 3-D face recognition. In the proposed algorithm, we first register the 3-D point clouds to provide a dense correspondence between faces. Afterwards, we analyze two mapping techniques—the closest-point mapping and the ray-casting mapping, to construct depth images from the corresponding well-registered point clouds. The depth images that are obtained are then divided into local regions where the discrete cosine transformation is performed to extract local information. The local features are combined at the feature level for classification. Experimental results on the FRGC version 2.0 face database show that the proposed algorithm performs superior to the well-known face recognition algorithms.

**Index Terms**—Automatic registration, depth image, local appearance face recognition, 3-D face recognition.

#### I. INTRODUCTION

Biometric identification is a challenging task that has received a significant amount of interest in the last decades. Among the utilized biometric modalities, the human face is one of the most natural. Moreover, a subject's face images can be acquired easily and unobtrusively. Due to low cost and the wide availability of image acquisition systems, most of the face recognition algorithms are based on 2-D intensity images [23]. However, the algorithms that process intensity images suffer from facial appearance variations that are caused by changes in head pose and illumination conditions. Much effort has been devoted to solving these problems in the 2-D domain. Although significant enhancements have been achieved in the 2-D domain against these variations under controlled conditions, the problem still remains unsolved under uncontrolled, real-world conditions.

Manuscript received May 1, 2007. This work was supported in part by the European Union under the integrated project CHIL, Computers in the Human Interaction Loop under Contract 506909, and in part by the German Research Foundation (DFG) as part of the Collaborative Research Center 588 Humanoid Robots-Learning and Cooperating Multimodal Robots. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bir Bhanu.

The authors are with the Department of Computer Science, University of Karlsruhe, Karlsruhe 76131, Germany (e-mail: kawagao1979@hotmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2007.902924

To handle the pose and illumination variations, utilizing 3-D shape information has been shown to be a promising approach [4] since a point cloud or surface can represent the geometric structure of the face, and is not affected by pose variations and by extrinsic sources of variations, such as illumination. A large number of approaches have been proposed for 3-D face recognition. Gordon [11] used depth and curvature information to describe face-shape features. Principal curvatures and directions are used to extract the location of the nose and eyes and the direction of the face. Principal direction-based enhanced Gaussian images were used by Tanaka *et al.* to represent facial surfaces [20]. Another powerful shape descriptor called point signatures was proposed by Chua *et al.* in [5], which is claimed to be invariant to facial expression changes. However, high computational cost is required to compute the point signature for each point. In [15] and [16], Lu *et al.* integrated the iterative closest point (ICP)-based rigid matching with the non-linear thin-plate spline (TPS) deformation. Feature vectors extracted from displacement vector field after deformation were classified with support vector machines (SVMs) [22]. The decision was made by combining the rigid matching distance and the deformation classification result. Irfanoğlu *et al.* also used ICP alignment to estimate landmarks automatically, and faces were recognized with the point set distance technique [14]. Although experiments in [3] show that TPS-based registration may have side effects in terms of discrimination, TPS for intrasubject and intersubject nonrigid deformations may increase performance in the case of expression variations [16]. Another approach for 3-D face recognition is to construct depth images from the registered point clouds and then to apply 2-D face recognition algorithms, such as eigenfaces or Fisherfaces [21], [24]. For a detailed recent survey of 3-D face recognition, please see [4].

In this paper, we present a novel 3-D face recognition algorithm that is based on local appearance face recognition. In the proposed algorithm, we first register the input point cloud in order to provide dense correspondence between the faces. Afterwards, we analyze two mapping techniques—closest-point mapping and ray-casting mapping to construct the depth images from the corresponding well-registered point clouds. Finally, we perform local appearance face recognition on these depth images. The local-appearance-based face recognition algorithm is proposed as a fast and generic approach [8], [9] and does not require the detection of any salient local regions. It partitions an aligned face image into nonoverlapping blocks of  $8 \times 8$  pixel resolution. The reason for having  $8 \times 8$  pixel block size is to provide sufficient compactness on the one hand, and to keep stationarity within the block on the other hand. The underlying ideas for preferring a local appearance-based approach over a holistic appearance-based approach are as follows: 1) In a holistic appearance-based face recognition approach, a change in a local region can affect the entire feature representation, whereas in local appearance-based face recognition, it affects only the features that are extracted from the corresponding block while the features that are extracted from the other blocks remain unaffected. This property provides robustness against both local registration imperfections and expression variations and 2) a local appearance-based algorithm can facilitate weighting of local regions. It can assign higher weights to regions which are found to be more discriminant.

In order to represent the local regions, the discrete cosine transform (DCT) is used. Its compact representation ability is superior to that of the other widely used input-independent transforms such as the Walsh–Hadamard transform. Although the Karhunen–Loève transform (KLT) is known to be the optimal transform in terms of information packing, its data dependent nature makes it infeasible for some practical tasks. Furthermore, DCT closely approximates the compact repre-

sentation ability of the KLT, which makes it very useful for representation both in terms of information packing and in terms of computational complexity.

The remainder of this paper is organized as follows. In Section II, we explain 3-D face shape registration and depth image generation techniques. We introduce the local appearance-based 3-D face recognition algorithm in Section III. Experimental results are presented and discussed in Section IV. Finally, in Section V, conclusions are given.

## II. FACE REGISTRATION AND DEPTH IMAGE CONSTRUCTION

Recorded point clouds may have different poses and expressions. In order to extract proper local information from corresponding local facial blocks, a precise point-to-point correspondence should be established. In the following subsections, the processing steps of the face registration and depth image generation are explained.

### A. Preprocessing

Range data acquired by 3-D sensors may be noisy and sometimes may have spikes with sharp disparity discontinuities on the surface. To remove spike artifacts, we applied a median filter. Afterwards, we used a Gaussian filter to make the face surface smoother. 3-D laser scanners may sometimes have difficulties imaging wet surfaces, such as eyeballs, and hairy surfaces, such as eyebrows, which may cause holes on the face surfaces. These holes were filled using linear interpolation.

### B. Dense Correspondence

To establish a dense correspondence between faces in training and testing sets, we transformed all faces to a common coordinate framework. This transform is based on landmarks that are placed on salient facial feature points. We first selected the face with the smallest number of points in the training set as a base mesh. Then, we placed 11 landmarks on all faces. We can use any set of landmarks as the common frame of reference, but in order to apply statistical shape analysis, such as the point distribution model (PDM) [6], we used the generalized Procrustes algorithm [12] to compute the mean landmarks. Each face is then warped onto the mean landmarks using the thin-plate spline transform. After aligning all of the surfaces, a point-to-point dense correspondence was established by selecting the surface's closest points to the vertices in the base mesh. Some faces in the data set may contain neck and ears while others may not, and since we are only interested in the face itself, the parts outside the face region should be removed. We discarded the vertices of the base mesh with a distance of more than 20 mm to the surfaces; for details, see [13]. The remaining vertices then construct the final base mesh. Aligned faces that are sampled with this base mesh will contain the same number of vertices and the same portion of the face region.

### C. Depth Image Generation

Resampling with closest-point mapping may result in folds and uneven sampling of the surface where the correspondence between surfaces with high curvature is not very close. In such case, as shown in Fig. 1(a), the tip of the target surface is not sampled at all, while the vicinity of that tip may be sampled twice, which introduces a fold into the final mesh. An example of a depth image generated from such a mesh is shown in Fig. 1(c). Pixels around the nose do not correspond to the exact depth value on the original face, which would degrade the recognition performance. Since the base mesh and target mesh are closely aligned, we used another resampling method illustrated in Fig. 1(b). Through each vertex on the base mesh, we cast a ray along the  $z$ -axis onto the target surface. The resampled point is the crossing point if it exists. However, vertices at the border of the base mesh may sometimes have no corresponding point. Then, the closest-point mapping was again applied because curvatures at the

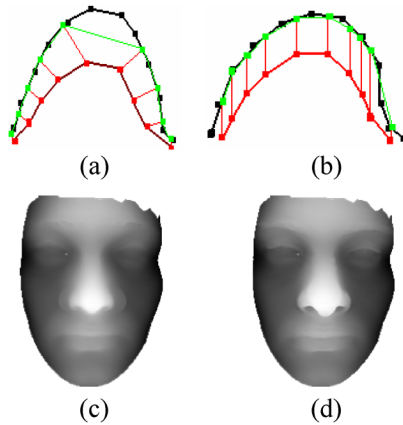


Fig. 1. (a) Closest-point mapping, find the closest point on the target mesh for each base mesh vertex. (b) Ray-casting mapping, find the crossing point on target mesh for each ray-casting from the base mesh vertices. (c) Depth map constructed with closest-point mapping. (d) Depth map constructed with ray-casting mapping.

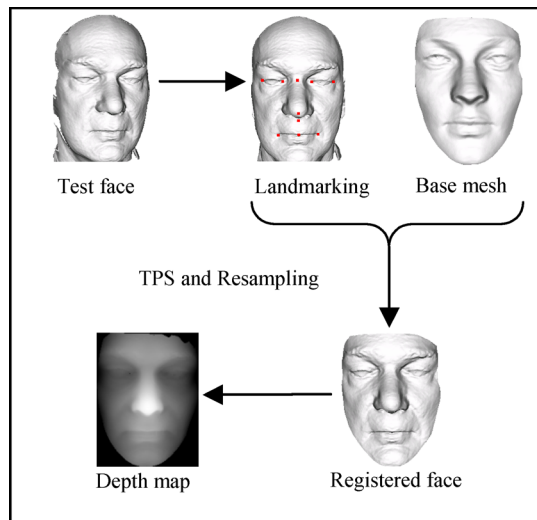


Fig. 2. Face registration and depth map image generation.

face mesh border are usually low. Depth map images generated with ray-casting mapping are better in appearance as shown in Fig. 1(d).

After resampling, the  $z$ -value of each vertex on the reconstructed face can be considered as an intensity value in the corresponding depth image. Fig. 2 illustrates the overall process of face registration and depth image generation from a 3-D image.

#### D. Automatic Registration

To construct a fully automatic recognition system, we need to register the face shapes automatically without interrupting the user. ICP-based registration is one of the most popular approaches for this purpose [4]. However, ICP easily falls easily into local minima if two shapes are not at least coarsely matched. Therefore, an initial alignment was performed by matching nose tips on two faces. The nose tip on each face was estimated according to the depth values and geometric information around the nose. Since the range images in the face recognition grand challenge (FRGC) version 2.0 data set [19] are close to frontal, we were able to coarsely transform the test face to the base mesh using the nose tip as an anchor point. But for range images with larger pose variations, one needs to use more anchor points for coarse matching. For instance, in [15], three points are used for coarse alignment.

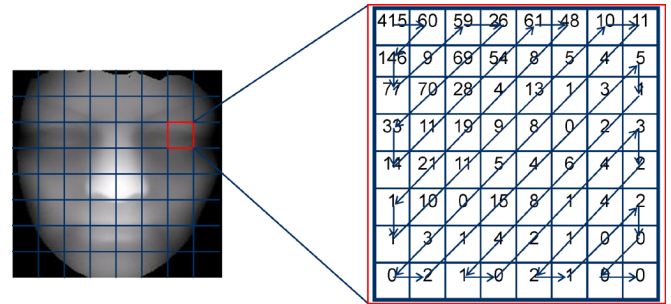


Fig. 3. Local blocks in depth image, DCT features are extracted using zig-zag scan.

ICP-based approaches treat the face as a rigid object, which is not suitable for handling expression variations [4]. Using a nonrigid deformation, such as the TPS technique, can deform faces with non-neutral expressions. Experiments in [16] show a substantial reduction of recognition errors after the nonrigid deformation. To perform TPS-based automatic registration, we need to detect the landmarks automatically. The automatic landmark detection algorithm we use is a modified version of the one in [14]. First, we compute surface normals, Gaussian curvatures, and mean curvatures for all vertices in the test face surface. The nose tip on the test face is estimated according to the depth value and the pose-invariant geometric information, and coarse matching is made with this estimated nose tip. After coarse matching, we perform ICP to align the test face to the base mesh more precisely so that we can estimate the initial positions of the landmarks on the test face according to the landmarks on the base mesh. Based on these initial estimates of the landmarks and the symmetry plane of the base mesh, the symmetry plane of the test face can be computed. Finally, the initial landmarks are fine tuned according to their surface normals, curvatures, and relative distance to symmetry plane.

### III. DISCRETE COSINE TRANSFORM-BASED LOCAL APPEARANCE MODELS

Local appearance face recognition is based on statistical representations of the nonoverlapping local facial regions and their combination at the feature level. The underlying idea is to utilize local information while preserving spatial relationships. In [9], the discrete cosine transform (DCT) is proposed to be used to represent the local regions. It has been shown to be a better representation method for modeling the local facial appearance compared to principal component analysis (PCA) and the discrete wavelet transform (DWT) in terms of face recognition performance.

Feature extraction from depth images using local appearance-based face representation can be summarized as follows: The input depth image is divided into blocks of  $8 \times 8$  pixels size. Each block is then represented by its DCT coefficients. These DCT coefficients are ordered using the zig-zag scanning pattern [10] (see Fig. 3). From the ordered coefficients,  $M$  of them are selected according to the feature selection strategy, resulting in an  $M$ -dimensional local feature vector. Finally, the DCT coefficients extracted from each block are concatenated to construct the overall feature vector of the corresponding depth image.

In order to compare the introduced local DCT-based representation with the depth representation, we calculated the ratio of within class variance to between class variance with each representation on a training set which has also been used for identification experiments. We calculated the ratio of within class variance to between class variance for each representation unit and then averaged it over the representation units and the subjects. We obtained an average ratio of

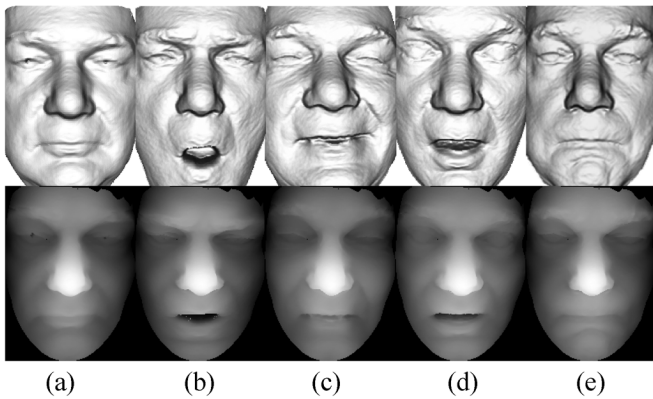


Fig. 4. First row: Preprocessed range images rendered with a shade model in training and test set. Second row: registered depth images. (a) Neutral. (b) Frowning. (c) Smiling. (d) Surprised. (e) Puffy.

0.5 with DCT-based local representation, and 0.67 with the depth representation. The lower ratio of within class variance to between class variance obtained by the proposed representation scheme indicates its better discrimination capability compared to the depth representation.

#### IV. EXPERIMENTS

We conducted extensive experiments on the FRGC version 2.0 data set [19] to analyze the performance of the proposed local appearance-based 3-D face recognition approach. The 3-D data corpus of the FRGC database was collected by imaging subjects with a range scanner. For our experiments, we selected the subjects who have at least two range images in the Spring 2003 recordings of the database, and used their images from these recordings for training. For testing, we used the range images of these subjects from the Spring 2004 recordings. The training data contain neutral expressions, whereas the testing data contain different expressions, such as frowning, smiling, etc. In total, we used 218 range images of 109 subjects for training, where each individual has two samples, and 758 range images for testing, where each individual has a different number of samples, ranging from one to 12. Sample pre-processed range images and the corresponding registered depth images from the training and testing datasets are shown in Fig. 4. The depth images are scaled to a resolution of  $64 \times 64$  pixels.

In the experiments, we used a nearest neighbor classifier with the L1 norm as distance metric, since it has been shown that the L1 norm provides better results than the L2 norm and normalized correlation [8]. We also tested an SVM classifier as a more sophisticated classifier.

##### A. Analysis of Local Appearance-Based 3-D Face Recognition

In the first part of the experiments, we analyzed the effects of local feature dimension, feature selection, and face registration on the face recognition performance. Fig. 5 shows the face recognition performance with respect to increasing local feature dimensionality. In this experiment, the 3-D faces were registered using all of the 11 manually labelled landmark points, and depth images were generated using ray-casting. At each local block, the first coefficient was removed from the ordered DCT coefficients, since it only represents the average depth of a local image block. From the remaining coefficients, the first  $M$  were selected. The selected local feature vector was normalized to have a unit norm as suggested in [8], which has been shown to improve the face recognition performance. As can be observed from the figure, high correct recognition rates can be attained by using only five-dimensional local feature vectors. The performance continues to increase slightly until the feature dimension of ten. The correct recognition rate remains the same or decreases slightly, when the

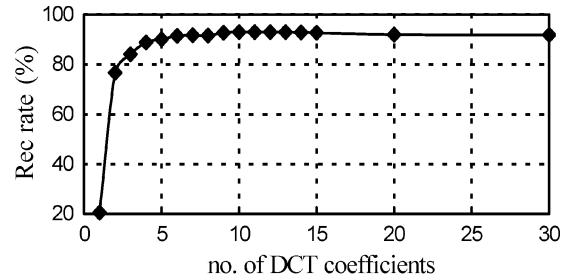


Fig. 5. Correct recognition rate versus local feature dimensionality.

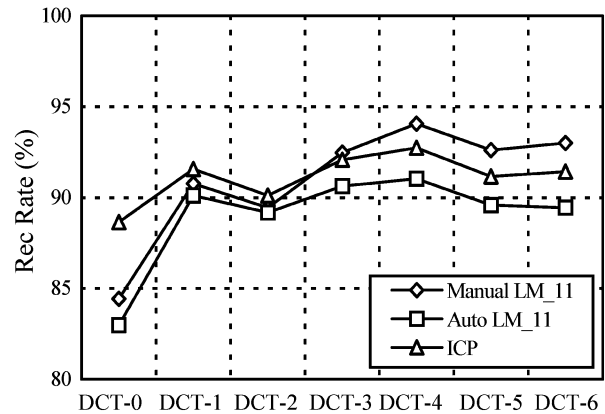


Fig. 6. Recognition rate of DCT-based local appearance approach using different feature sets. (DCT- $N$ : Discard first  $N$  coefficients and select the first 10 coefficients from the remaining ones.)

dimensionality increases further. Therefore, we chose to use ten dimensional local feature vectors for the rest of the experiments.

The second experiment assesses the effect of frequency content on the face recognition performance. In order to consider the correct recognition rate with different sets of features having different frequency contents, we discarded the first  $N$  ( $N = 0, 1, \dots, 6$ ) low-frequency DCT coefficients and conducted the face recognition experiments. We ran the same experiments with three different registration setups to observe whether the selected features produce consistent results over each registration framework. The registration configurations were named “Manual LM\_11,” “Auto LM\_11,” and “ICP.” “Manual LM\_11” and “Auto LM\_11” correspond to face registration with 11 manually and automatically labeled landmarks, respectively, whereas “ICP” corresponds to registration with ICP. Ray-casting was used to generate depth images from the registered 3-D faces. Correct recognition rates obtained from these three different experimental setups are plotted in Fig. 6. In all of the experiments, the best results were obtained using the DCT-4 feature set, which implies that removing the coefficients that represent horizontal and vertical changes as well as the one that represents the average depth, improves the face recognition performance.

The effects of the landmark points used for registration and the depth image generation techniques were analyzed in the third experiment. Usually using more landmarks for registration improves correspondence, but if the landmark points are poorly placed, correspondence may get worse. If more landmark points than necessary are used while performing the TPS warping, the cumulative noise of the landmarks may result in degenerate deformations. Therefore, we discarded some of the landmarks to analyze their effectiveness. In the experiment, five possible landmark combinations, illustrated in Fig. 7(a), were tested for registration. Both ray-casting and closest-point methods were used for depth image generation. The DCT-4 feature set was used

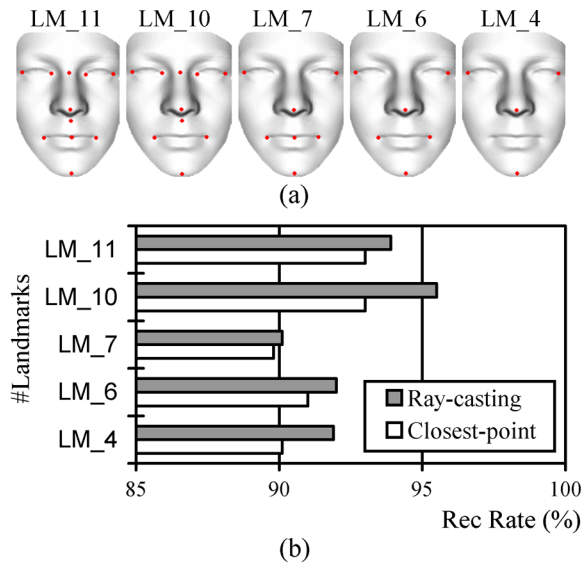


Fig. 7. (a) Five landmark combinations. (b) Recognition rate of a DCT-based local appearance approach with different landmark combinations.

TABLE I  
MANUAL REGISTRATION VERSUS AUTOMATIC REGISTRATION

Registration method	Recognition Rate
Manual LM_10	95.5%
ICP	92.7%
Automatic LM_10	93.1%

for classification. The corresponding results can be seen in Fig. 7(b). We achieved the highest scores by selecting ten landmarks, excluding the landmark located in the middle of the mouth. This is expected since this point is not easy to label on faces that have different facial expressions. As can be observed, ray-casting mapping always outperforms closest-point mapping. With optimal landmark combination and ray-casting, we achieved a 95.5% correct recognition rate.

We investigated the effect of automatic landmarking in the last experiment. Table I compares the performance of the proposed face recognition algorithm on the 3-D face images that are registered using manually labeled landmark points, automatically with ICP and using automatically detected landmark points. The depth images were generated by ray-casting and the DCT-4 feature set was used for classification. The results obtained using the automatic registration methods—with ICP and with automatically detected landmark points—are slightly lower than the results obtained on the images that are registered using the manual labels. This decrease in performance is mainly caused by the errors introduced by ICP registration and automatic landmark detection. Better results were attained on the images that are registered using automatically detected landmark points than on the ones registered via ICP. This indicates that deformation onto a common frame is able to mitigate the effects of expression variations.

### B. Analysis of Different Bases

In this part of the experiment, we compared the DCT-based local appearance representation with different well-known basis functions that can also be used for representing the local regions. In addition, we also compare the proposed local appearance-based approach with the discrete cosine transform-based holistic approach. In these experiments, the 3-D faces were registered using ten manually labelled landmark points, and depth images were generated using ray-casting.

TABLE II  
PERFORMANCE COMPARISON OF LOCAL APPEARANCE REPRESENTATION METHODS

Method	Performance
DCT	95.5%
KLT	84.0%
WHT	92.4%
FT	80.6%
WT	74.5%

TABLE III  
LOCAL DCT VERSUS HOLISTIC DCT

Method	Performance
Local DCT	95.5%
Holistic DCT	84.0%

We compared the DCT with the KLT, Walsh–Hadamard transform (WHT), Fourier transform (FT), and wavelet transform (WT). For the KLT, we used 20-D local feature vectors; for the WHT, we used the same feature setup as the one we used for DCT; for the FFT, we used the magnitudes of the Fourier coefficients. For the WT, we used the Daubechies 4 wavelet, which has been shown to perform better in terms of computation time and recognition performance with respect to the other order Daubechies wavelets and other well-known wavelets [7]. We used the first-order scaling component as the feature vector [7]. Table II gives the correct recognition rates obtained with each basis function. As can be seen, the DCT achieved the best result. After the DCT, the WHT also reaches a high correct recognition rate compared to the other basis functions. The other basis functions attained lower performance although they used higher dimensional feature vectors. The feature dimension was 20 in the KLT, 64 in the FT, and 16 in the WT, whereas it was 10 in the DCT and the WHT.

The comparison of DCT-based local appearance and holistic 3-D face recognition is given in Table III. In the holistic approach, the same dimensional feature vector is selected for the entire image using the same feature selection strategy as the one used for the local appearance-based approach, that is, removing the first four DCT coefficients and selecting the remaining first 640 DCT coefficients that are ordered according to the zig-zag pattern. The results show the importance of applying DCT locally and then combining the local analysis results in order to construct the overall feature vector.

The results from Tables II and III indicate that the obtained performance improvement with the proposed algorithm is not solely based on conducting classification in the frequency domain or using the DCT. For instance, the Fourier transform-based local appearance and the DCT-based holistic 3-D face recognition approaches have been found to perform poorly. This shows that the performance improvement is provided by performing local analysis and using the DCT for representing the local regions.

### C. Performance Comparison

In the final part of the experiments, we compared the proposed local appearance-based 3-D face recognition approach with several well-known face recognition algorithms: Eigenfaces [21], linear discriminant analysis (LDA) [24], Bayesian face recognition [17], local binary patterns (LBP) [2], embedded hidden Markov model (EHMM) [18], point set difference (PSD) [14], and point distribution model (PDM) [6]. For the local appearance-based approach, we also used an SVM classifier instead of a nearest neighbor classifier to assess the performance of a more sophisticated classification scheme.

Table IV shows the experimental results of each algorithm. The correct recognition rates attained on manually and automatically

TABLE IV  
PERFORMANCE COMPARISON OF METHODS WITH MANUAL  
AND AUTOMATIC LANDMARK-BASED REGISTRATION

Method	Manual LM 10	Automatic LM 10
Local DCT	95.5%	93.1%
Local DCT+SVM	90.0%	89.0%
LBP	91.7%	90.5%
EHMM	87.9%	85.5%
Eigenfaces	88.6%	86.5%
LDA	92.4%	88.5%
Bayesian	94.9%	89.7%
PSD	81.4%	80.6%
PDM	87.6%	84.7%

registered images are given. In eigenfaces, Bayesian face recognition and PDM algorithms, we used 100 principal components. This is the number of principal components with which we achieved the best results. For LDA, we used the LDA+PCA algorithm provided in the CSU face identification evaluation system [1]. This version of LDA uses a soft-distance measure proposed by Zhao *et al.* [24]. For LBP, we used the  $LBP_{8,2}^{u_2}$  operator—uniform patterns in a circular (8,2) neighborhood—which is the operator used in [2]. In EHMM, we used a  $4 \times 4$ -size DCT coefficients matrix as an HMM observation, which is extracted from a  $12 \times 12$  image block by applying the DCT. In local appearance 3-D face recognition, both for nearest neighbor and for SVM classification, we used the 10-D DCT-4 feature set. The radial basis function was the kernel function in the SVM classifier.

From the results given in Table IV, it can be observed that the proposed local appearance-based approach outperforms the other well-known face recognition algorithms as well as the local DCT features classified with SVM, which may suffer from a small training set problem. The performance of all algorithms decreases slightly when they use the images that are registered using automatically detected landmarks. These results indicate that the proposed local DCT features provide a powerful and robust representation of the depth images for classification purposes.

## V. CONCLUSION

In this paper, we proposed a depth image-based 3-D face recognition approach using local appearance-based models. Depth images were obtained by a base mesh-based registration and a resampling technique. We extracted the local features from each block on a depth image using the DCT, and then concatenated the local features in order to conserve spatial information.

We conducted extensive experiments on the range images from the FRGC version 2.0 face database to investigate several factors that may affect the recognition performance. First, we performed face recognition experiments using different local feature dimensionality. We observed that the correct recognition rate increases at the beginning with growing feature vector dimensionality. It reaches the best result with a 10-D local feature. After this point, the recognition rate remains the same or decreases slightly. The second experiment investigated different feature sets and frequency contents. We found that the DCT-4 feature set, which excludes the horizontal and vertical depth changes, as well as the average depth value, attains the best results. Third, we analyzed landmark selection for TPS-based registration to improve dense correspondence. In the experiments, we obtained the best result by discarding the landmark located on the middle of the mouth. During these experiments, we also observed that ray-casting mapping always outperforms closest-point mapping when face surfaces are resampled with the base mesh. In the last experiment, we assessed the performance

of the fully automatic systems. We compared two automatic registration methods—rigid ICP and nonlinear TPS warping based on automatic landmarking. Face recognition performance decreased slightly with automatic registration. However, having only a slight reduction in the correct recognition rate due to automatic registration indicated that it is possible to have an online, fully automatic 3-D face recognition system without sacrificing much performance. We also evaluated other well-known basis functions in addition to the DCT for local appearance representation. We found that DCT performs significantly better than the other basis functions. The Walsh–Hadamard transform obtained the second best result. In addition, we compared the DCT-based local appearance and holistic approaches. We observed that it is very important to perform the DCT locally. Finally, we thoroughly compared the proposed local appearance-based approach with well-known face recognition algorithms (PCA [21], LDA [24], LBP [2], PDM [6], PSD [14], Bayesian [17], EHMM [18]) as well as with the local appearance-based approach using the SVM classifier. Experimental results showed that the proposed algorithm provides an improvement over existing algorithms in face recognition performance.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their useful suggestions. The authors would also like to thank M. Fischer, J. Stallkamp, and J. McDonough for their contributions to the study.

## REFERENCES

- [1] The CSU Face Identification Evaluation System [Online]. Available: <http://www.cs.colostate.edu/evalfacerec/>.
- [2] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
- [3] L. Akarun, B. Gökberk, and A. A. Salah, "3D face recognition for biometric applications," presented at the 13th Eur. Signal Processing Conf., Antalya, Turkey, 2005.
- [4] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Comput. Vis. Image Understanding*, vol. 101, pp. 1–15, 2006.
- [5] C. S. Chua, F. Han, and Y. K. Ho, "3D human face recognition using point signatures," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, 2000, pp. 233–237.
- [6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Comput. Vis. Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [7] H. K. Ekenel and B. Sankur, "Multiresolution face recognition," *Image Vis. Comput.*, vol. 23, no. 5, pp. 469–477, 2005.
- [8] H. K. Ekenel and R. Stiefelhagen, "Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization," presented at the CVPR Biometrics Workshop, New York, 2006.
- [9] H. K. Ekenel and R. Stiefelhagen, "Local appearance-based face recognition using discrete cosine transform," presented at the 13th Eur. Signal Processing Conf., Antalya, Turkey, 2005.
- [10] R. C. Gonzales and R. E. Woods, *Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, 2001.
- [11] G. Gordon, "Face recognition based on depth and curvature features," in *Proc. IEEE CVPR*, 1992, pp. 108–110.
- [12] J. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.
- [13] T. Hutton, B. Buxton, and P. Hammond, "Dense surface point distribution models of the human face," in *Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, HI, 2001, pp. 153–160.
- [14] M. O. İrfanoğlu, B. Gökberk, and L. Akarun, "3D shape-based face recognition using automatically registered facial surfaces," in *Proc. ICPR*, 2004, vol. 4, pp. 183–186.
- [15] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–43, Jan. 2006.
- [16] X. Lu and A. K. Jain, "Deformation analysis for 3D face matching," presented at the IEEE WACV, Breckenridge, CO, Jun. 2005.

- [17] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognit.*, vol. 33, no. 11, pp. 1771–1782, Nov. 2000.
- [18] A. Nefian, "A hidden Markov model-based approach for face detection and recognition," Ph.D. dissertation, Dept. Elect. Comput. Eng. Elect. Eng., Georgia Inst. Technol., Atlanta, 1999.
- [19] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," presented at the IEEE CVPR, San Diego, CA, Jun. 2005.
- [20] H. T. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-based face surface recognition using spherical correlation-principal direction for curved object recognition," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, 1998, pp. 372–377.
- [21] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Sci.*, pp. 71–86, 1991.
- [22] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [23] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: A literature survey," *ACM Comput. Surveys*, vol. 35, no. 44, pp. 399–458, 2003.
- [24] W. Zhao, R. Chellappa, and P. J. Phillips, "Subspace linear discriminant analysis for face recognition," UMD TR4009, 1999.

## Face Verification Using Template Matching

Anil Kumar Sao and B. Yegnanarayana

**Abstract**—Human faces are similar in structure with minor differences from person to person. These minor differences may average out while trying to synthesize the face image of a given person, or while building a model of face image in automatic face recognition. In this paper, we propose a template-matching approach for face verification, which neither synthesizes the face image nor builds a model of the face image. Template matching is performed using an edginess-based representation of the face image. The edginess-based representation of face images is computed using 1-D processing of images. An approach is proposed based on autoassociative neural network models to verify the identity of a person. The issues of pose and illumination in face verification are addressed.

**Index Terms**—Autoassociative neural network (AANN), face verification, 1-D image processing.

### I. INTRODUCTION

The objective of the face verification task is to accept or reject the identity claim of the person using his or her face image [1]. The issues involved in this task can be categorized into two classes, namely, 1) interclass variation and 2) intraclass variation. The interclass variation refers to the differences in the face images of two people, which are due to uniqueness of the features present in the face image of each person. The intraclass variation refers to the differences in the face images of a given person under varying conditions of pose, illumination, and expressions [1].

Template matching is one of the approaches proposed in the literature to address the issue of interclass variation [2], because it takes

Manuscript received November 1, 2006; revised May 10, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rama Chellappa.

A. K. Sao is with the Department of Computer Science and Engineering, Indian Institute of Technology-Madras, Chennai 600 036, India (e-mail: anil@cs.iitm.ernet.in).

B. Yegnanarayana is with the International Institute of Information Technology, Hyderabad 500 032, India (e-mail: yegna@iiit.ac.in).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2007.902920

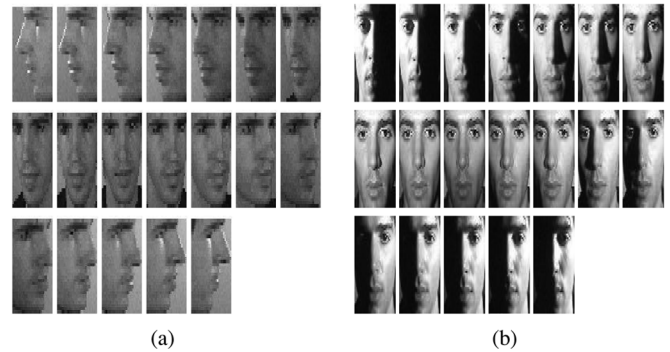


Fig. 1. Face images of a person with (a) pose variation and (b) illumination variation.

the unique information of a person's face image into account. But this approach has the drawback that it gives poor performance under intraclass variation [3]. The problem of intraclass variation can be overcome using an approach which synthesizes a 3-D model of the face image from a given sample [4]–[8]. But the synthesis of face image may result in some artifacts and some loss of the unique information. Thus, the synthesis-based approach may degrade the performance of the face verification system. Another way to address the issue of intraclass variation is to consider several reference face images which capture variations in the face images, such as different poses or due to different lighting conditions. These reference face images can be used to build a model for that person's face image. The model is used to verify the identity of a test face image. Such methods are discussed in [9]–[14]. In these cases, the model may average out some of the information that is unique for that person.

In this paper, we propose a template-matching-based approach, which neither synthesizes the face image nor derives a model for the person's face. We use reference face images (at different poses or at different lighting conditions) separately for template matching. The template matching is performed using an edginess-based representation of a face image [15]. The scores obtained by template matching with different reference images are combined in a selective way. The combined scores are used for verification by using the autoassociative-neural-network (AANN) model-based classifier.

The performance of the proposed approach is evaluated on the FacePix database collected at Arizona State University [16], [17]. The FacePix database consists of 30 people, each having two sets of face images: A set with pose-angle variation, and a set with illumination angle variation. The set with pose-angle variations has 181 images (representing angles from  $-90^\circ$  to  $90^\circ$  at  $1^\circ$  interval). In this paper, we denote these images by  $I^1, \dots, I^{181}$ . The illumination set is captured with the subject looking directly into the camera while the light source is moved around the subject. The light source moves at a  $1^\circ$  interval from  $-90^\circ$  to  $90^\circ$ . These images are denoted by  $L^1, \dots, L^{181}$ . Some of the face images of a person are shown in Fig. 1. In our experiments, the size of the face images is rescaled to  $30 \times 30$  pixels.

The organization of this paper is as follows. Section II explains the template matching using the edginess-based representation of a face image. The scores obtained by matching several templates are combined in a selective way as explained in Section III. An approach is proposed in Section IV to classify the combined scores using AANN models. Experimental results are discussed in Section V, and a summary of the work is given in Section VI.