Robust 3D Face Recognition Using Learned Visual Codebook

Cheng Zhong, Zhenan Sun and Tieniu Tan
National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, 100080, P.R.China
{czhong,znsun,tnt}@nlpr.ia.ac.cn

Abstract

In this paper, we propose a novel learned visual codebook (LVC) for 3D face recognition. In our method, we first extract intrinsic discriminative information embedded in 3D faces using Gabor filters, then K-means clustering is adopted to learn the centers from the filter response vectors. We construct LVC by these learned centers. Finally we represent 3D faces based on LVC and achieve recognition using a nearest neighbor (NN) classifier. The novelty of this paper comes from 1) We first apply textons based methods into 3D face recognition; 2) We encompass the efficiency of Gabor features for face recognition and the robustness of texton strategy for texture classification simultaneously. Our experiments are based on two challenging databases, CASIA 3D face database and FRGC2.0 3D face database. Experimental results show LVC performs better than many commonly used methods.

1. Introduction

Automatic identification of human faces is a very challenging research topic, which has gained much attention during the last few years. Most of this work, however, is focused on intensity or color images of faces [1]. There is a commonly accepted claim that face recognition in 3D is superior to 2D because of its invariance to illumination and pose variations. Recently with the development of 3D acquisition system, 3D face recognition has attracted more and more interest and a great deal of research effort has been devoted to this topic.

Many methods have been proposed for 3D face recognition over the last two decades [2]. Some earlier research on curvature analysis has been proposed for face recognition, which can characterize delicate features [3]; Chua et al. [4] treat face recognition as a 3D non-rigid surface matching problem and divide the human face into rigid and non-rigid regions. The rigid parts are represented by point-signatures to identify the individual. Beumier et al. [5] develop a 3D acquisition prototype based on structure light and build a 3D face database. They also propose two methods of surface matching and central/lateral profiles to compare two instances. However, these methods usually require a high computation cost and involve small databases. In addition, because of the sensitivity of curvature based features, these methods also need high quality 3D face data. Bronstein et al. [6] propose a method capable of extracting the intrinsic geometric features of facial surfaces using geometric invariants, and the use of bending invariant canonical representation makes it robust to facial expressions and transformations typical of nonrigid objects. Lu et al. [7] construct many 3D models as registered templates, then they match 2.5D images to these models using iterative closet point (ICP). Chang et al. [8] use principal component analysis (PCA) on both 2D intensity images and 3D depth images, and fuse 2D and 3D results to obtain the final performance. Their results show that appearance based methods such as PCA can also give a good performance for 3D face recognition.

In this paper, we introduce a novel LVC for 3D face recognition. The flowchart is shown in Fig.1. Our method can be divided into learning section and recognition section. For both sections, because data noise and expression variations are the main obstacles for 3D face recognition, we preprocess the raw 3D data and extract the face area which is robust to expression variations. To reduce the texture complexity, we also divide 3D face into many local patches. We calculate the Gabor features for each patch and filter response vectors are stored into its corresponding 3rd-order tensor. In learning section K-means clustering is adopted to learn center vectors from each tensor and these centers are stored into the corresponding sub-codebook. LVC is constructed by concatenating all of these sub-codebooks. In recognition section, we obtain some histograms between each local patch and its corresponding sub-codebook, and the representation of each 3D face is composed by concatenating all such histograms. We achieve face recognition using NN classifier finally.

The contributions of this paper are as follows. First, we propose a novel LVC method to extract the intrinsic dis-

1-4244-1180-7/07/$25.00 ©2007 IEEE

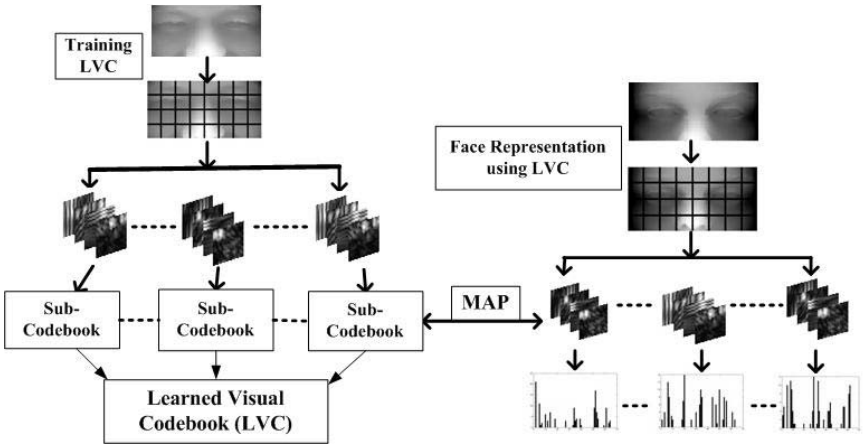
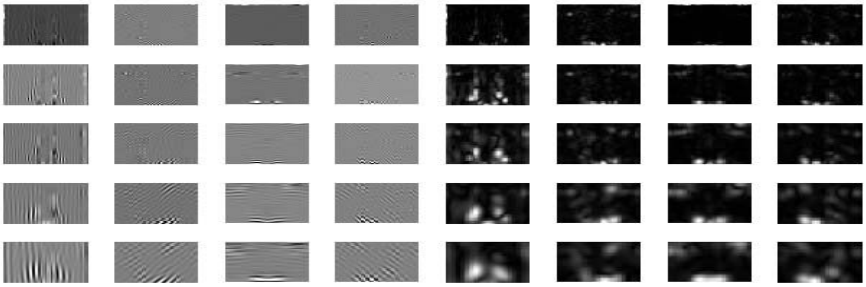
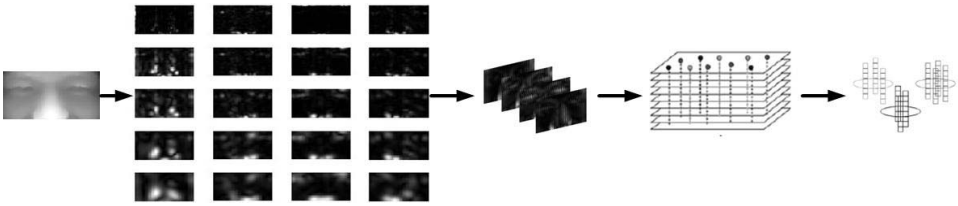


Figure 1. The flowchart of LVC scheme.



(a) 3D Gabor Faces. Left is real part of the representation and right is the magnitude part of the representation.



(b) Convertion from a 3D face into many basic facial elements, the center of each cluster is corresponding to basic facial element.

Figure 2. Gabor features of 3D face.

criminative information from 3D faces for recognition. Second, we encompass the effectiveness of gabor features for face recognition and the robustness of texton strategy for texture classification together. Third, we make a detailed comparison between LVC and some commonly used appearance based methods, such as PCA, gabor features (GF) and local binary pattern (LBP) [9][10][11]. Experimental results show the effectiveness of LVC to characterize 3D faces for recognition.

The remainer of this paper is organized as follows. In Section 2, we give our motivation of LVC. In Section 3, we introduce how to represent 3D faces and achieve recognition using LVC. We describe our experimental results in Section 4. Finally, the paper is concluded in Section 5.

2. Motivation

Recently, many researchers pay attention to textons based methods for texture classification. Because of its statistical characteristics, this method is suitable to recognize different objects in images and has been successfully applied into visual recognition tasks [12][13]. Motivated by

this, we suppose that each 3D face is composed of many basic facial elements. Our main task is to learn these elements from the given 3D faces and construct a visual codebook to encompass all such elements.

To achieve this target, we also need to find effective features to extract intrinsic discriminative information embedded in 3D faces. There are many choices to satisfy our request, such as LBP, ordinal filters and so on [11][14]. Because of the excellent performance on spatial locality and orientation selectivity [10], in this paper we choose Gabor features to represent 3D faces. In this way, we encompass the efficiency of Gabor features for face recognition and the robustness of texton strategy for texture classification simultaneously.

3. 3D face representation and recognition

3.1. Preprocessing of 3D faces

There are two kinds of 3D face data in our experiments, CASIA and FRGC2.0 [15]. For CASIA data (WRL format), first we automatically detect the nose-tip point based on effective energy. Then a trained SVM classifier is adopted to determine whether it is a nose-tip point [16][17]. When nose-tip point is located, we construct many mesh models corresponding to 3D faces. The correspondence of these mesh models is determined by the position of nose-tip, then ICP is applied to complete the registration [18]. For FRGC2.0 data (ABS format) [15], we use manually labeled points to complete the registration. Finally, for both kinds of registered data, we crop a 150*120 region from the raw 3D data to construct a 3D face image which centered at nose section. From Fig.3 we find that mouth area is very sensitive to the variations of expressions (such as laugh and surprise), therefore, we only use the 3D faces without mouth area in our experiments and the size of each image is 80*120.

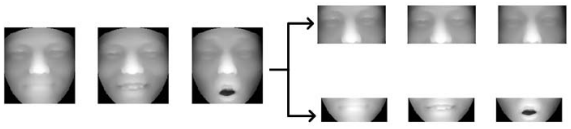
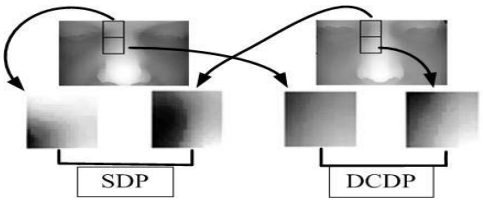


Figure 3. 3D faces with different expressions.

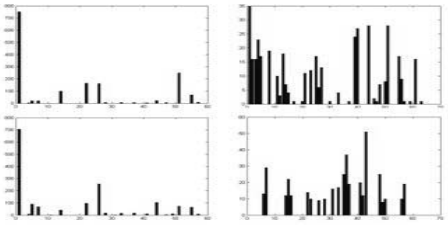
3.2. 3D Gabor Face

Texture is often characterized by its responses to a set of orientation and spatial-frequency selective linear filters. Because of the excellent performance on orientation and spatial frequency selectivity, in our method we use Gabor filters to extract intrinsic discriminative information embeded in 3D faces [10]. The Gabor filters can be defined by Eq.1:

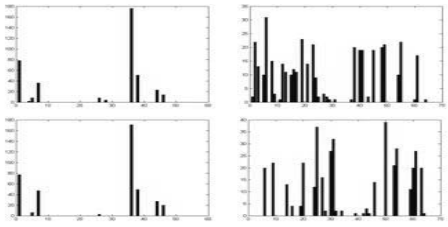
$$\Psi_{u,v}(z) = \frac{||k_{u,v}||^2}{\sigma^2} e^{(-\frac{||k_{u,v}||^2||z||^2}{2\sigma^2})}[e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}}] \quad (1)$$



(a) A comparison of local patches from different people. SDP means patches with different facial structures, DCDP means patches with different depth-contrast.



(b) Left is the LBP representation of SDP, right is the LVC representation of SDP.



(c) Left is the LBP representation of DCDP, right is the LVC representation of DCDP.

Figure 4. A comparison between LBP and LVC based on local patches from different people.

where $z = (x, y)$ and u and v denote the orientation and scale of Gabor kernels, respectively. Here we use Gabor kernels with 5 scales and 4 orientations. The number of scales and orientations is selected to represent the facial characteristics of spatial locality and orientation selectivity.

The Gabor representation of 3D face image, called 3D Gabor Face, is the convolution of the image with the Gabor kernels. From Fig.2(a) we can find that most of the energy is distributed near the nose section, which means that nose section contains most of the discriminative information for 3D face recognition. Finally each 3D Gabor Face is stored in a 3rd-order tensor, as Fig.2(b) shown.

3.3. Local patches strategy

3D face contains many basic facial elements and variations based on these elements. Therefore, it is a difficult work to construct such a huge visual codebook to cover all these elements and variations. To solve this problem, we divide the 3D face into some local texture patches and construct a sub-codebook in correspondence with each patch. In this way, we not only reduce texture complexity but also add some spatial information into these local patches.

In addition, different from 2D face images, for local patch croped at the same location in 3D face images, different people contain different basic facial elements, as Fig.4(a) shows. Therefore, we can also increase difference between different subjects and reduce difference between the same subject.

The size of local patch will influence the recognition result. If the size is small, there is much data noise. On the other side, if the size is large, we will lose a lot of spatial information. In our experiment we choose 20*20 as local patch size, which is a tradeoff between texture statistical information and spatial information.

3.4. Learned Visual Codebook

LBP is an efficient texture descriptor and has been successfully used into face recognition [11]. But the patterns in LBP are predefined before classification, which makes them not the best descriptors for a specific application, such as 3D face recognition. To overcome this problem, we need to learn these patterns from the given training subjects.

The effectiveness of Gabor features for face representation has been proved in many works [10]. However, it should be noted that we want to characterize an individual. We do not expect the Gabor filter response vectors to be totally different at each pixel over the images. Thus, there should be several center vectors and all others are noisy variations of them. Based on this hypothesis, we use K-means method to cluster these Gabor filter response vectors into a small set of center vectors, as Fig.2(b) shows [13]. We can view these centers as the learned patterns from training 3D face images, which are the basic facial elements. LVC is composed by all these learned facial elements.

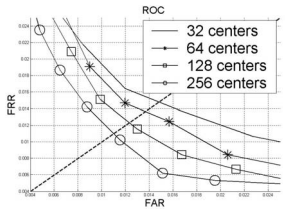
We carry out some experiments to test the capability of LBP and LVC to represent 3D faces. As Fig.4(a) shows, we choose two kinds of patches from different subjects. One pair is the patches with different facial structures, which is called as SDP. The other pair is the patches with the same structure but different depth-contrast information, which is called DCDP. From Fig.4(b), we find that two LBP histograms have only a little difference to represent SDP, while the difference from two LVC histograms is remarkable. The situation is more explicit in Fig.4(c). The two LBP histograms are almost the same to represent DCDP, while we can also find remarkable difference between two LVC histograms. In addition, we find that histogram distribution covers most bins in LVP representation, while histogram distribution in LBP representation only covers a few bins. Experimental results show that the performance of LVP to characterize 3D faces is substantially better than LBP.

From definition of LVC, we find that assigning different number of centers for each local patch will influence the recognition performance. We make a comparison based on FRGC1.0 database, as Fig.5(a) shows. From the experiment we find that the recognition performance is improved little by assigning more centers for each patch. To reduce computation cost, in our experiment we choose 64 centers to represent each patch in 3D faces.

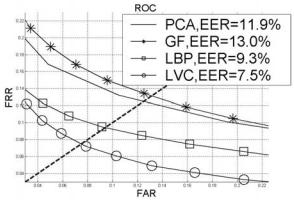
3.5. 3D face representation and recognition

In our experiment, the size of each 3D face image is 80*120 and we choose 100 images as the training set. If we directly store these vectors into a 3rd-order tensor, the size of the tensor is 5*4*80*120*100 and it contains 960000 20-dimensional vectors. It is a huge number for clustering. To solve this problem, we also need to divide each image into some local patches, and assign each local patch a corresponding 3rd-order tensor to store their filter response vectors. Therefore, our LVC method can be summarized as follows:

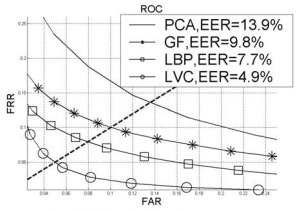
- Learning section.
  - We divide 3D face images into lots of local patches and Gabor filters are adopted. The filter response vectors of each patch are stored into its corresponding 3rd-order tensor.
  - K-means clustering is adopted on each tensor and we obtain many sub-codebooks using these learned centers.
  - We construct LVC by concatenating all sub-codebooks together.
- Recognition section.
  - We also divide 3D face images into lots of local patches and map Gabor filter response vectors of each patch into their corresponding sub-codebooks in LVC.
  - A mapping vector between each 3D face and LVC is constructed by concatenating the corresponding histograms between each local patch and its corresponding sub-codebook, which is our representation of 3D face characteristics.
  - We achieve face recognition using NN classifier. Here we use L_1 as distance measure.



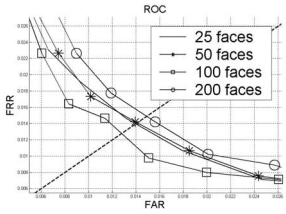
(a) Recognition performance when assigning different number of centers for each local patch.



(b) ROC performance on CASIA Database.



(c) ROC performance on FRGC Database.



(d) Recognition performance when using different number of training images.

Figure 5. ROC performances in our experiments.

3.6. Other methods used for comparison

PCA, GF and LBP are commonly used appearance based methods in face recognition and all of these methods achieve very good performance [9][10][11]. Here we make detailed comparisons between LVC and the above three methods to show the efficiency of our proposed method for 3D face recognition.

4. Experimental results and discussion

Our proposed method LVC is evaluated in terms of its representation and recognition capacity on two very challenging databases, CASIA 3D face database and FRGC2.0 3D Face Database [15].

Figure 6. Some example images in our experiments.

4.1. Experiments on CASIA 3D Face Database

There are 123 persons in CASIA 3D Face Database, which contains variations of illuminations, expressions and poses. In our experiment we use 15 images for each person, which contain 5 images with neutral expression and 10 images with different expressions (smile, laugh, anger, surprise and eye closed), and we have 1845 3D face images in total. Some images from CASIA 3D Face Database are shown in the first row of Fig.6. Because some methods in our experiment, such as PCA and LVC, need learning section, we choose the first 100 images in this database as training set, and the total of the 1845 images as testing set. LVC method is compared with some widely used methods PCA, GF and LBP, experimental results are shown in Fig.5(b).

4.2. Experiments on FRGC 2.0 3D Face Database

FRGC2.0 3D Face Database is the most challenging database as far as we know. In this database it contains variations of sessions, expressions, illuminations and so on, which make it more challenging for 3D face recognition. Some images from FRGC2.0 3D Face Database are shown in the second row of Fig.6. In this experiment we don't follow the rules as FRGC2.0, which use the 943 images in FRGC1.0 as training set and the left 4007 3D face images as testing set. Instead of that, we only choose the first 100 3D faces from this database as training set, and use all of the FRGC2.0 database, 4950 3D face images in total, as our testing set. Our proposed LVC method is compared with some widely used methods PCA, GF and LBP and experimental results are shown in Fig.5(c).

4.3. Discussion

From experimental results we can find that: First, representing 3D face texture using LVC achieves the best performance for 3D face recognition on both experiments, which shows the effectiveness of our method to describe the characteristics of 3D faces.

Second, as Fig.5(b) shows, PCA method performs better than GF method. As far as we know, GF is more efficient features for face recognition and it should give a better performance than PCA. The main reason for this result is that we registrate 3D faces in CASIA database automatically and there are some registration errors in 3D face data. Because GF can extract more delicate features embedded in 3D faces, its recognition performance will be influenced greater by these errors than PCA. However, this sensitivity problem of GF is solved by our LVC method, in respect that it is a statistical representation of 3D faces and more robust to such small errors.

Third, LVC not only represents statistical information but also mines some specific intrinsic discriminative information embeded in 3D faces. According to definition, LBP can be recognized as a special case of textons based method. The main difference between it and our LVC method is that the textons in LBP, such as $LBP_{(8,1)}$, is predefined in terms of statistical importance. However, LVC can extract textons

according to specific application. Because of the K-means clustering learning section, we can tune our textons choice according to different textures, which can be considered as adaptive textons selection.

We also find that learning section in LVC does not need many training images. We make an experiment to compare the recognition performance by using different number of training images, which is based on FRGC1.0. As Fig.5(d) shows, the recognition result changes from EER 1.5% to EER 1% . The notable point is that we only use a small number of training images to obtain a good recognition result, which means LVC only need a small size of training set to learn how to effectively describe 3D face characteristics.

4.4. Future work

In our experiments, many recognition errors are due to registration errors in both databases, which reflects the importance of preprocessing for 3D face recognition. Many domains need to be done to improve our preprocessing, such as nose location, face registration and hole filling.

5. Conclusion

In this paper, we have proposed a novel LVC for 3D face recognition. In our method we not only extract intrinsic discriminative information embedded in 3D faces using Gabor features, but also choose K-means clustering to learn basic facial elements and construct a Learned Visual Codebook. In this way we encompass the efficiency of Gabor features for face recognition and the robustness of texton strategy for texture classification simultaneously. Experimental results show that the performance of LVC is better than other commonly used methods.

6. Acknowledgement

This work is funded by research grants from the National Basic Research Program (Grant No.2004CB318110), Natural Science Foundation of China (Grant No.60335010, 60121302, 60275003, 60332010, 69825105, 60605008), Hi-Tech Research and Development Program of China (Grant No.2006AA01Z193) and the Chinese Academy of Sciences.

References

[1] W. Zhao, R. Chellappa, and A. Rosenfeld. Face recognition: a literature survey. *ACM Computing Surveys*, 35:399–458, 2003.

[2] K. W. Bowyer, K. I. Chang, and P. J. Flynn. A survey of approaches to three-dimensional face recognition. *International Conference on Pattern Recognition*, 1:358–361, 2004.

[3] G. Gordon. Face recognition based on depth maps and surface curvature. *SPIE*, pages 234–274, 1991.

[4] C. Chua, F. Han, and Y. Ho. 3d human face recognition using point signature. *IEEE Inetrnational Conference on Automatic Face and Gesture Recognition*, pages 233–238, 2000.

[5] C. Beumier and M. Acheroy. Face verification from 3d and grey level cues. *Pattern Recognition Letters*, 22:1321–1329, 2001.

[6] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 2005.

[7] X. Lu, A. K. Jain, and D. Colbry. Matching 2.5d face scans to 3d models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):31–36, 2006.

[8] K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multi-modal 2d+3d face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, 2005.

[9] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, (1), 1991.

[10] Chengjun Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11(4):467–476, 2002.

[11] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. *European Conference on Computer Vision*, pages 469–481, 2004.

[12] F. Jurie and B. Triggs. Creating efficient codebooks for visual recognition. *IEEE International Conference on Computer Vision*, 1:604–610, 2005.

[13] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[14] Z. Sun, T. Tan, and Y. Wang. Robust encoding of local ordinal measures: A general framework of iris recognition. *ECCV workshop on Biometric Authentication*, pages 270–282, 2004.

[15] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[16] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification, Second Edition*. John Wiley & Sons, Inc., 2001.

[17] C. Xu, Y. Wang, T. Tan, and L. Quan. A robust method for detecting nose on 3d point cloud. *IEEE International Conference on Image Processing*, 2004.

[18] C. Xu, Y. Wang, T. Tan, and L. Quan. Automatic 3d face recognition combining global geometric features with local shape variation information. *IEEE International Conference on Automatic Face and gesture Recognition*, 2004.