

A Method Based on Moment Invariants and Sparse Representation for 3D Face Recognition

Chao Zhang^a and Yanning Zhang^{a,*}

^aShaanxi Provincial Key Laboratory of Speech & Image Information Processing
School of Computer Science, Northwestern Polytechnical University,
Xi'an, 710129, China

Abstract. This paper presents a moment invariants based three-dimensional face recognition method under sparse representation framework. An improved moment invariants descriptor is constructed and the discriminative ability in face recognition is validated on NPU3D and FRAV database. A series of partial faces are obtained by segmenting each face model according to the geodesic distance from nose middle point. Proposed moments invariants is computed and stacked as the feature of each face model. Finally, sparse representation is used in the classification step. The result on GavabDB shows that this method is very promising compared with ICP baseline.

Keywords: moment invariants; sparse representation; 3-D face recognition; NPU3D face database

1. Introduction

With the rapid development of three-dimensional information acquisition technique and device, it is possible for us to recognize the shapes of 3-D objects directly, without need to map the 3-D shape on 2-D domain and change the 3-D recognizing problem to image-based task.

Face recognition is a burgeon research field in recent years, there are many different kinds of recognizing methods, especially using 3D face information. Due to the limitation of pages, the details will not be included in this paper.

2-D moment invariants were firstly proposed by Hu [1] in 1962 in the application of character recognition. Sadjadi and Hall [2] first extended moment invariants from 2-D to 3-D. Lo and Don [3] constructed 3-D moment invariants with complex moments and group-theoretic technique. Moments of surface patch was also mentioned in [4]. File format such as .wrl and .off are commonly used representations of 3-D models, which include coordinates of vertexes and polygonal patches. Li [5] used Gaussian theorem to convert a volume integral into a surface one and decreased the computational complexity of polyhedral moments. JM Pozo et al [6] proposed an efficiently algorithm to exactly compute the geometric moments on surface via pre-compute the trinomial offline. Moments can also be approximately computed by accumulating the Cartesian coordinated of the volume pixels sampled in the body of the objects. In Xu et al [7], 3-D surface moment defined on the surface of the 3-D models was proposed and 6 surface moment invariants under similarity transformation are given.

Sparse representation, also known as compressed sensing, has been applied to image-based face recognition and showed promising future. Under this framework, each face is represented by a set of features, which include sufficient information of the individual. In the statistical signal processing community, the algorithm problem of computing sparse linear representation with respect to an overcomplete dictionary of base elements has seen a recent surge of interest. The excitement centers around the discovery that whenever

* Corresponding author. Tel.: +86-13060399678; fax: +86-02988431536.
E-mail address: ynzhang@nwpu.edu.cn.

the optimal representation is sufficiently sparse, it can be efficiently computed via convex optimization. In the work of Wright et al [8], the discriminative nature of sparse representation in classification was exploited. They represent the test sample in an overcomplete dictionary whose base elements are the training samples themselves. Motivated by [8], in order to be robust to expressions, in the work of Li and Jia [9], after a uniform remeshing procedure, they construct features which are insensitive to severe expressions for each face and finally apply sparse representation framework in classification step.

Most experiments are conducted on the famous public face database such as GavabDB [10] and FRGC v2.0. However, one problem is that these databases only include foreign people, so it is not sufficient to prove the effectiveness in Chinese database. Our experiments are conducted both on the famous foreign database and on the Chinese database. The Chinese database we use the NPU3D database collected by Northwestern Polytechnical University. For more information about the database, you can refer to our official website: <http://www.saiip-npu3d.org>.

This paper presents an approach to solve 3D face recognition task, at the same time to reduce the computational complexity and running time. We extract the moments invariant descriptor mentioned in [7] as feature, and apply the sparse representation technique to classify the test data. Before the recognition step, all the data should be preprocessed in order to sweep the negative effect brought by the acquisition step using 3D laser scanner. The 3D models we use are collected under normal illumination condition using Konica Minolta 910 laser scanner.

2. Method

We characterize each face by a descriptor, whose elements are a set of moment invariants. Before the calculation of the moment invariants, pre-processing is necessary to eliminate the noise data of the surface and to ensure the correctness of the result. Then, we finally recognize 3D faces under the framework of sparse representation using the stacked moment invariants.

2.1. Pre-processing

The first step is geometric processing. Although the nose tips can easily detected through choosing the vertex that has the largest z-axis depth, we find that there are still several faces which have the wrong nose tips due to the noise around them or the pose variation. We use Meshlab [11] to adjust the pose of the face surface and eliminate the unwanted parts such as hair and collar. Other geometric processing operations such as smoothing and hole-filling are also done by Meshlab. The comparison of face before and after the operation can be seen in Figure 1.



Fig. 1 Pre-processing sample (a) before; (b) after

2.2. Partial Surface Segmenting

Region selecting means to select a part from the processing surface, as we know, the salient feature of face mainly center around the nose tip. But using only one single region, the dimension of the features is too few to be enough discriminative. So, we selected a series partial face using different radius from the anchor points such as nose middle point in our method. The range of the radius is also a problem need to determine. In our experiment, we choose the radius from 40mm to 80mm and achieve good result. The illustration of the region selecting can be seen in Figure 2. The nose tip is determined by choosing the vertex having the maximum depth, and then fast marching on triangulated domains (FMTD) [12] algorithm is used to compute the geodesic distance from arbitrary start point.



Fig. 2 Selected partial surface using different radii. Top: samples centering on nose tip; Down: samples centering around nose middle point.

2.3. Moment Invariants Description

After we get the multi-level region of the face surface, we can forward to compute the moment invariants for them.

Three-dimensional surface moments of order $l + m + n$ are defined by the surface integrals defined on the surface area S of $P(u, v)$:

$$M_{lmn} = \iint x^l y^m z^n \rho(x, y, z) ds \quad (1)$$

Where $\rho(x, y, z)$ is the density function defined on the surface. As to face surface, we can set it equal to 1.

The moment invariants we use to describe the face are proposed by Xu [7]. We first calculate the geometric moments on surface up to order N . And then the proposed moment invariants are computed using the equation given in [7].

Apart from the 6 moment invariants proposed in [7], we consider using three 2-order moments as feature. High order moments are easily affected by the noise and bring out error in describing face surface, however, low order moments are more robust to noise and more reliable in describing objects.

2.4. Sparse Representation Classification

The final step is to classify a face surface to the right identity it belong to. Under the sparse representation framework, test probe y is represented by a linear combination of training samples. Each sample v can be described by a column vector of dimension m , suppose the number of training set A is n , then the test sample can be represented like:

$$y = Ax_0 \in R^m, \quad (2)$$

where

$$A = [v_{1,1}, v_{1,2}, \dots, v_{1,n_1}, v_{2,1}, \dots, v_{k,n_k}] \in \mathfrak{R}^{m \times n} \quad (3)$$

is the training set, and

$$x_0 = [0, \dots, 0, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}, 0, \dots, 0]^T \in \mathfrak{R}^n \quad (4)$$

is a coefficient vector whose entries are mostly zero except those associated with the i -th subject.

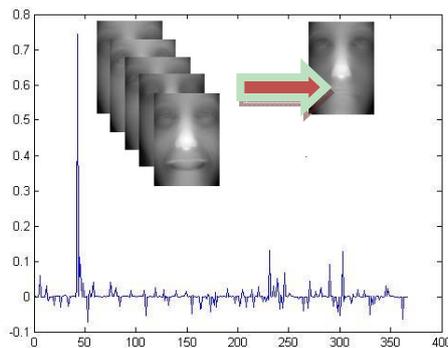


Fig. 3 Test sample can be reconstructed from the linear combination of training samples belonging to the same class

We use the reduced dimension sparse representation. Applying R to both sides of equation yields:

$$\tilde{y} = RAx_0 \in \mathfrak{R}^d \quad (5)$$

where $R \in \mathfrak{R}^{d \times m}$, $d \ll m$.

Recent development in the emerging compressed sensing theory reveals that if the solution is sparse enough, the combinatorial problem is equivalent to the following l^1 -minimization problem:

$$\min \|x\|_1 \quad \text{subject to } \tilde{y} = \tilde{A}x. \quad (6)$$

After the optimization procedure, we can compute the residuals; the identity is determined by choosing the one having the smallest residual.

To ensure higher stability by aggregating the results from multiple random projections, we use the following algorithm. When the projection matrix R is randomly generated, we seek a robust estimate of the residuals r_i by using an ensemble of random face matrices $R_j, j = 1, \dots, l$. Define r_i^j as the residual of the test sample belonging to i th class using the projection matrix R^j , and then the empirical average of the i -th residual over l projection is

$$E[r_i] = \frac{1}{l} \sum_{j=1, \dots, l} r_i^j \quad (7)$$

Then the identity y can be assigned as

$$ID(y) = \arg \min_i E(r_i) \quad (8)$$

3. Experiments

3.1. Experiment on NPU3D and FRAV database

The goal of this experiment is to validate the discriminative ability of moment invariants in classification. We use the database with little expression and noise. The program is running on Matlab2009. The first experiment we conducted is on NPU3D and FRAV database.

NPU3D database contains 300 people and each person has 35 scans with different pose, expression and occlusion. In order to test the effectiveness of our method, we only use a subset of the database which includes 50 person and each with 3 scans. The sample can be seen in figure 4.



Fig.. 4 Samples in NPU3D database

The number of male and female is both 25. In the test, we divide the samples into three groups and run the recognition program three times. Each time, we select one group as test set, and the other two as training set. The classifier we use is simply the nearest neighbor classifier. We compare the result of two different surface representations. Surface form and isomeric surface mentioned in [15].The overall recognition result can be seen in the table 1 below:

Table 1. Result on NPU3D

Surface type	Identity	Scans	Total	Rate
Original	50	3	150	.8519
Embeding	50	3	150	.4833

Table 2. Result on FRAV

Surface type	Identity	Scans	Total	Rate
Original	20	8	160	.7619
Embeding	20	8	160	.7250

From the result, we can see that the moment invariants on surface are reliable features in the frontal face recognition. As mentioned in the paper [8], increasing the measurement can get rather good result. The stacking of the moment invariant gives birth to the sparsity of the facial feature.

3.2. Experiment on Gavab DB

The goal of the second experiment is to prove the effectiveness and efficiency of our method under sparse representation framework. The GavabDB database is the international public three-dimensional face database. The database includes 549 scans of 61 identity (45 is male, 16 is female). The data is purely surface shape data, without corresponding texture. Each identity has 9 scans of different pose and expression. We only use 7 scans of them because we want to test the method on pose-varied and expression-varied dataset.

We use the preprocessing method mentioned above to smooth and crop the face surface. After the regions are selected and successive moment invariants are computed as the feature descriptor under the sparse representation framework.

In the experiment, the dataset is divided into 7 groups, the first group is pose-varied group and another 3 group is expression-varied group. Although the samples in pose-varied group have been registered to uniform coordinate space, there are many samples suffering data missing problem. In order to handle this problem, we need to repair the face surface and fill the missing data using minimal surface. To some extent, this operation is beneficial to the step of selecting region which is based on the geodetic distance on the face surface.

In the expression-varied group, the geometric deformation in the mouth area has changed the topology structure of the facial surface largely, thus increasing the difficulty to realizing satisfactory classification. Here, in order to improve the discriminative ability of moment invariants descriptor, the origin of the geodesic distance is set at the nose middle point.

We evaluate the performance of two kinds of mask in the region selecting process. One is called nose tip centering strategy (NT); another is called nose middle point centering strategy (NM). The difference between them is the choice of the origin. We expect the later strategy could lead to more robustness to the expression changing.

As to the construction of the feature vector, we compare the moments invariants(MI6) given in [7] and an improved version MI9 of that, we add three second order moment invariants to the MI6 to form a more discriminative descriptor which performs better under sparse representation framework.

The test result can be seen in the table below:

Table 3. Experiment result using two kinds of masks and two kinds of feature vectors

	<i>NF1</i>	<i>NF2</i>	<i>PV1</i>	<i>PV2</i>	<i>EV1</i>	<i>EV2</i>	<i>EV3</i>
NM+MI6	.87	.77	.80	.80	.67	.71	.60
NM+MI9	.90	.87	.83	.83	.77	.80	.67
NT+MI6	.90	.87	.83	.80	.67	.70	.50
NT+MI9	.90	.90	.83	.70	.57	.60	.53
ICP	.92	.92	.74	.74	.62	.62	.62

NT stands for Nose Tip method, NM stands for Nose Middle point method

Table 4. Experiment result using Randomface

	<i>NF1</i>	<i>NF2</i>	<i>PV1</i>	<i>PV2</i>	<i>EV1</i>	<i>EV2</i>	<i>EV3</i>
NM+MI6	.87	.73	.80	.80	.73	.70	.60
NM+MI9	.90	.87	.83	.80	.77	.77	.60
NT+MI6	.87	.90	.80	.70	.57	.60	.53
NT+MI9	.90	.87	.83	.80	.63	.67	.47

The random matrix is $m \times m$, dimension reduction is not made to give the best performance

In table 3, the first 2 columns belong to frontal group, the identification rate is about 90%; column 3 and 4 belong to pose-varied group, after the registering operation and preprocessing operation, the recognition result is above that 80%; the last three columns belong to the expression-varied group, in our method, we just rely

on the nose middle point segmenting to handle the expression effect, obviously, it is far from enough. However, our method is not designed to be robust to expression, so the degrading result can be expected.

Compared with ICP baseline, the identification rate is nearly the same, and the performance on pose-varied group and expression-varied group is better than ICP, in addition, our method has advantage in reducing running time due to avoiding of iterative surface matching and registering process.

From the comparison of the two tables above, no significant difference is observed which means that randomface still catches the importance facial information without losing too much accuracy.

4. Conclusion

In this paper, we present a novel three-dimensional recognition method using moment invariants under sparse representation framework. We also give some experiments results and comparison to validate the effectiveness and efficiency of the proposed method.

In the first experiment, we give the performance of moment invariants in classification task using frontal facial database. The result shows that moment invariants is a good choice comparing to other surface matching based method because of the rotation invariant property. So, it can be used as feature descriptor in recognizing three-dimensional faces.

Traditional classifiers such as NN classifier and nearest subspace classifier have their limitation in recognizing facial faces especially in three-dimensional face recognition problem. Motivated by the recent related work, we propose to use sparse representation to solve this challenging problem.

In addition, more improvement can be made in the future as follows.

In our experiments, we find that moment invariants descriptor is sensitive to non-rigid deformation such as changing expression, although some isomeric embedding method have been developed, the features extracted from the deformed surface is not reliable and easily affected by noise.

In the preprocessing step, we use the built-in smooth algorithm of MeshLab to smooth the facial surface, if the smooth iteration or parameter is not proper set, the dissimilarity between different face models is also diminished after smoothing. A feature-preserved smoothing method is useful in our experiment to improve the discriminative ability of the moment invariants.

In the sparse representation, random matrix is applied to the original data to reduce the number of measurement. It is not clear whether the selection of the moment invariants or the number of it will improve the performance of proposed method.

5. Acknowledgements

The authors would like to acknowledge the use of the GavabDB face database in this paper due to Moreno and Sanchez [10]. This work is partly supported by NWPU start foundation (Z2010073).

6. References

- [1] M. K. Hu, "Visual Pattern Recognition by Moment Invariants", *IRE Trans. Information Theory*, 1962, vol. 8, pp.179-187.
- [2] F. A. Sadjadi and E. L. Hall, "Three-Dimensional Moment Invariants", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1980, vol. 2, no. 2, pp. 127-136.
- [3] C. H. Lo and H. S. Don, "3-D Moment Forms: Their Construction and Application to Object Identification and Positioning", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1989, vol. 11, no. 10, pp. 103-1064.
- [4] C. H. Lo and H. S. Don, "Pattern Recognition using 3-D Moments", *Proc. 10th Int. Conf. Pattern Recognition*, 1990, pp. 540-544.
- [5] B. C. Li, "The Moment Calculation of Polyhedra", *Pattern Recognition*, 1993, vol. 26, no. 8, pp. 1229-1233.
- [6] J. M. Pozo, M. C. Villa-Urilo, A. F. Frangi, "Efficient 3D Geometric and Zernike Moments Computation from Unstructured Surface Meshes", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 09 Aug, 2010.
- [7] Dong Xu and Hua Li, "Geometric Moment Invariants", *Pattern Recognition*, 2008, vol. 41, no. 1, 240-249.

- [8] J. Wright, A. Yang, A. Ganesh, S. Sastry and Yi. Ma, “Robust Face Recognition via Sparse Representation”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2009, vol. 31, no. 2, pp. 210-227.
- [9] Xiaoxing Li, Tao Jia and Hao Zhang, “Expression-insensitive 3D Face Recognition using Sparse Representation”, *CVPR 2009*, pp. 2575-2582.
- [10] A. Moreno, and A. Sanchez. “GavabDB: A 3D Face Database”, *Workshop on Biometrics on the Internet COST275*. March 2004, pp. 77-85.
- [11] <http://meshlab.sourceforge.net/>
- [12] R. Kimmel, and J. A. Sethian, “Computing Geodesic Paths on Manifolds”, *Proc. Nat’l Academy of Science*, 1998, vol. 95, pp. 8431-8435.
- [13] A. Elad and R. Kimmel, “On Bending Invariant Signatures for Surfaces”, *IEEE. Trans. Pattern Analysis and Machine Intelligence*, 2003, vol. 25, no. 10.