3D HUMAN FACE RECOGNITION USING SUMMATION INVARIANTS

Wei-Yang Lin, Kin-Chung Wong, Nigel Boston, Yu Hen Hu

University of Wisconsin-Madison
Department of Electrical and Computer Engineering
1415 Engineering Drive, Madison WI, 53706 USA

ABSTRACT

A novel family of geometrically invariant features, called *summation invariants* are proposed for the recognition of the 3D surface of human faces. In particular, a 2D semi-local summation invariant feature is extracted from each column and each row of a rectangular region surrounding the nose of a 3D facial depth map. Through extensive experimentation, we empirically identify the most efficient 2D summation invariant features. We also investigate the proper pre-processing method for the 2D summation invariant features. Tested with the 3D facial data from the Face Recognition Grand Challenge v1.0 dataset, the proposed new features exhibit significant performance improvement over the baseline algorithm distributed with the dataset.

1. INTRODUCTION

Human face recognition has received unprecedented interest in recent years [1, 2]. However, rigorous tests with real-world data such as FERET, FRVT have revealed many shortcomings of existing approaches [3, 4]. In short, for large scale, real world situations, current systems still cannot deliver the performance needed for practical applications.

A majority of current face recognition approaches make use of 2D frontal facial texture features. Nonetheless, facial texture features are sensitive to lighting, pose, distance, age (temporal) variations, and can easily be altered through simple make-up efforts. On the other hand, from 3D facial surfaces, one may exploit features that are invariant to appearance variations. For example, the facial surface around cheek bones or the nose would remain unchanged under varying lighting conditions, are less likely to change due to aging, and are seldom covered with hairs. Hence, in this research, we focus on exploiting *invariant* features extracted from 3D facial surfaces.

Invariants for transformation groups play an important role in computer vision. Classical differential invariants such as curvature depend on derivatives that may be very sensitive to noise [5, 6]. Several approaches such as the semi-differential invariant introduced in [7, 8], integral invariant [9], and affine quasi-invariant arc-length [10] have been proposed. These invariants are defined on continuous functions. When applied to digitized object descriptions of contours or surfaces, numerical approximation will be needed and the results can be affected significantly by step size and other detailed settings.

Recently [11], we introduced a general method to generate invariants that are weighted summations of discrete data, as analogues to integral ones. Since these invariants are defined explicitly on discrete data, they do not require computationally intensive numerical integration to compute and will not be affected by the choice of step size. On the other hand, using weighted summation to compute the summation invariant will greatly reduce the impact of noise and hence promises higher signal to noise ratio for the computed invariant features. Specifically, in [11], we proposed a semi-local summation invariant feature for two-dimensional (2D) closed contours. It delivers superior performance compared to those produced using integral invariants or wavelet invariant features.

In this paper, we exploit the feasibility of applying 2D summation invariant feature for 3D facial range images classification. Several key design issues are addressed: (a) To identify, among many possible variations of 2D summation invariant features in the family, those features that yield highest performance. (b) To investigate the impacts of various preprocessing methods, including scaling and normalization on the proposed invariant features. (c) To explore proper feature reduction methods that will enhance the computation efficiency. In addition to analytical derivations, we have conducted extensive experiments using the Face Recognition Grand Challenge v1.0 dataset and the BEE (Biometric Experiment Environment) package. Compared to the best performance obtained using the FRGC [12] 3D baseline algorithm, the proposed features yield significant performance improvement.

The rest of this paper is organized as follows. Section 2 describes the summation invariant. Section 3 illustrates our algorithm in detail. In Section 4, we present the experimental results and compare them with those of the FRGC 3D baseline algorithm. Finally, Section 5 summarizes our contributions and provides an overview of future directions.

Wei-Yang Lin, Nigel Boston, and Yu Hen Hu are supported by the National Science Foundation under Grant No. CCF-0434355.

2. SUMMATION INVARIANTS

The method of moving frames [5, 6], originally introduced by Élie Cartan, is a powerful tool for constructing invariants under group actions. For expression-neutral face images, pose variations may be well modeled with Euclidean geometrical transformations. In this paper, we will use profile curves obtained by slicing 3D facial surfaces. Hence, we will focus on deriving novel invariants for the Euclidean group acting on $\mathbb{R}^2$. This is similar to our recent work deriving summation invariants for the affine transformation group [11].

Given a curve $(x[n], y[n])$ with $n = 0, 1, \ldots, N-1$, we use $(\bar{x}[n], \bar{y}[n])$ to denote this curve under Euclidean transformations.

$$\begin{bmatrix} \bar{x}[n] \\ \bar{y}[n] \end{bmatrix} = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} \begin{bmatrix} x[n] \\ y[n] \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix}$$

and we can find a moving frame by solving the following equations.

$$(\bar{x}[0], \bar{y}[0], \bar{y}[N-1]) = (0, 0, 0)$$

From the moving frame, a family of invariant functions can be derived. We define the invariant functions $\eta_{i,j}$ as below

$$\eta_{i,j} = \sum_{n=0}^{N-1} \bar{x}^i[n]\bar{y}^j[n]$$

where $i, j \in \mathbb{N}$. The first and second order invariant functions, i.e. $i + j = 1$ or 2, have been explicitly derived as shown below:

$$\begin{aligned}
\eta_{1,0} &= P_{1,0}(x_1 - x_0) + P_{0,1}(y_1 - y_0) \\
&\quad + Nx_0(x_0 - x_1) + Ny_0(y_0 - y_1) \\
\eta_{0,1} &= P_{1,0}(y_1 - y_0) + P_{0,1}(x_0 - x_1) \\
&\quad + N(x_1y_0 - x_0y_1) \\
\eta_{2,0} &= -2P_{1,0}(x_0 - x_1)(x_0^2 - x_0x_1 + y_0^2 - y_0y_1) \\
&\quad -2P_{0,1}(y_0 - y_1)(y_0^2 - y_0y_1 + x_0^2 - x_0x_1) \\
&\quad +P_{2,0}(x_0 - x_1)^2 + P_{0,2}(y_0 - y_1)^2 \\
&\quad +2P_{1,1}(x_0 - x_1)(y_0 - y_1) \\
&\quad +N(x_0(x_0 - x_1) + y_0(y_0 - y_1))^2 \\
\eta_{1,1} &= P_{1,1}((x_0 - x_1)^2 - (y_0 - y_1)^2) \\
&\quad +P_{1,0}(y_0^3 + 2x_0x_1y_1 - 2y_1x_0^2 \\
&\quad +x_0^2y_0 - 2y_0^2y_1 + y_0y_1^2 - x_1^2y_0) \\
&\quad -P_{0,1}(x_0^3 + 2y_0y_1x_1 - 2x_1y_0^2 \\
&\quad +y_0^2x_0 - 2x_0^2x_1 + x_0x_1^2 - y_1^2x_0) \\
&\quad +(P_{0,2} - P_{2,0})(x_0 - x_1)(y_0 - y_1) \\
&\quad +N(x_1y_0 - x_0y_1)(x_0(x_1 - x_0) + y_0(y_1 - y_0)) \\
\eta_{0,2} &= 2(x_1y_0 - x_0y_1)(P_{1,0}(y_1 - y_0) - P_{0,1}(x_1 - x_0)) \\
&\quad +P_{2,0}(y_0 - y_1)^2 + P_{0,2}(x_0 - x_1)^2 \\
&\quad -2P_{1,1}(x_0 - x_1)(y_0 - y_1) + N(x_0y_1 - x_1y_0)^2
\end{aligned}$$

where $x_0 = x[0]$, $x_1 = x[N-1]$, $y_0 = y[0]$, $y_1 = y[N-1]$, and

$$P_{i,j} = \sum_{n=0}^{N-1} x^i[n]y^j[n].$$

The summation invariant is defined over a segment of any 2D curve under Euclidean transformation group action. In order to enhance the discriminating power of this feature, we have [11] proposed a *semi-local* summation invariant feature that evaluates an invariant function for each pixel of the curve over a local window surrounding that pixel. As such, a curve consisting of N pixels will generate a feature vector of the same length.

3. APPLICATION TO 3D FACE RECOGNITION

3.1. 3D face dataset and BEE

We use the Face Recognition Grand Challenge (FRGC) dataset [12] to conduct face recognition experiments. FRGC is sponsored by the US National Institute of Standard and Technology (NIST) and other government agencies. Its testing data contains comprehensive 3D range images of human faces. It also provides an XML-based framework to document and describe computation experiments called the Biometric Experimentation Environment (BEE). All our experiments are conducted using BEE.

The 3D data provided by FRGC v1.0 contains 275 subjects (1 to 8 range scans per subject) and a total of 943 range scans. Each range scan has a resolution of 640×480 pixels. It contains both a 2D texture image and a 3D range data. In this paper, we use only the 3D range data in all experiments.

3.2. Procedures

FRGC defines four experiments, among which experiment 3 concerns the 3D face recognition task. Our experiment procedures follow closely what has been defined for the baseline algorithm provided by FRGC except for the following modifications. Details of experiment 3 are omitted here due to space limitation.

1. Use 3D data only: In order to focus only on 3D face recognition performance, the decision fusion part in experiment 3 is disabled so that 2D texture results are not included.

2. Specify the region of interest: Instead of using the whole 3D range face as the baseline algorithm does in BEE, we extract invariant features from an $N \times N$ ($N = 81$) rectangular region centered at the nose tip. An example is shown in Figure 1.

3. Alignment refinement: In FRGC, the location of the nose tip is manually selected. We use the provided nose

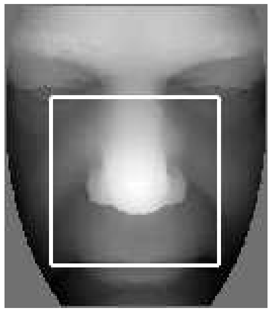

Fig. 1. Normalized depth map and an 81×81 region centered at the nose tip.

tip location to define an initial $N \times N$ region. The summation invariants computed from the $N \times N$ region is called the *summation image*. For all the normalized range data, we compute their summation images from the initial region and find their average image, called *the mean of summation images*. Then, for each normalized range data, we find a new $N \times N$ region which has minimal SSD (sum of squared differences) with the mean of summation images and compute a summation invariant from the new region. This procedure leads to more accurate alignment of range data.

4. Arc-length resampling : Range data are resampled uniformly with respect to arc-length. Specifically, for each row on a range data, we first compute its arc-length and resample it uniformly with respect to arc-length. Then, we perform the same resampling on each column.
5. Semi-local summation invariants: While the baseline algorithm uses 3D range data directly, we extract a semi-local summation invariant from each row and each column of the 81 $\times$ 81 rectangular region and use the results as invariant features. The length of the local window surrounding each pixel for calculating the semi-local invariant is chosen to be $L = 21$.

Then, we perform PCA on the resulting invariant features.

4. EXPERIMENTAL RESULTS

We have conducted a series of experiments to assess the performance of the proposed algorithm. In our experiments, we choose $N = 81, L = 21$ and Mahanolibis cosine as distance measure.

4.1. Effects of different summation invariants

In this experiment, we compare the ROC performance of different summation invariants. From the ROC curves in Fig 2, it is apparent that not all semi-local summation invariants are created equal in terms of discriminating power.

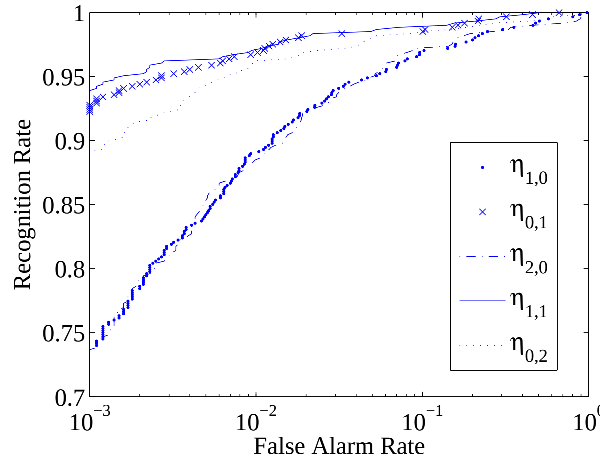

Fig. 2. ROC performance for different summation invariants.

4.2. Effects of difference scaling of x and y coordinates

Recall that the summation invariants developed in section 2 are for the Euclidean group acting on $\mathbb{R}^2$. While this feature is invariant to Euclidean transformations, it will be affected by improper scaling of coordinates. In the baseline algorithm, x and y coordinates in the range data are discarded. In our method, however, we need to retain these pieces of information. We experimented with different values of increments, $dx = dy \in \{0.4, 0.6, 0.8, 1.0\}$, and a fixed scaling factor 12.5 on z values. In Figure 3, we observe a significant impact of these changes on the ROC curve. In general, better recognition performance is expected if the original x, y, z information is preserved. In this experiment, only the summation invariant $\eta_{0,1}$ is used.

4.3. Comparison with FRGC 3D baseline algorithm

In this section, we conduct three experiments to evaluate the performance of our algorithm and the FRGC 3D baseline algorithm. In the first, we simply run the FRGC 3D baseline algorithm. In the second, we still run the FRGC 3D baseline algorithm but using only the cropped region rather than the whole normalized range data. In Fig 4, the second experiment shows a lower recognition rate than the first one. This is reasonable because the second experiment uses less data to perform recognition. In the third experiment, we use $\eta_{1,1}$ and apply our algorithm on the cropped region. Our algorithm yields the highest recognition rate as one can see in Fig 4. The results clearly indicate that summation invariants offer statistically significant better recognition performance than the range data itself. Note that these three experiments use exactly the same PCA parameters, dropping the first 10 eigenvectors and using the following 100 eigenvectors.

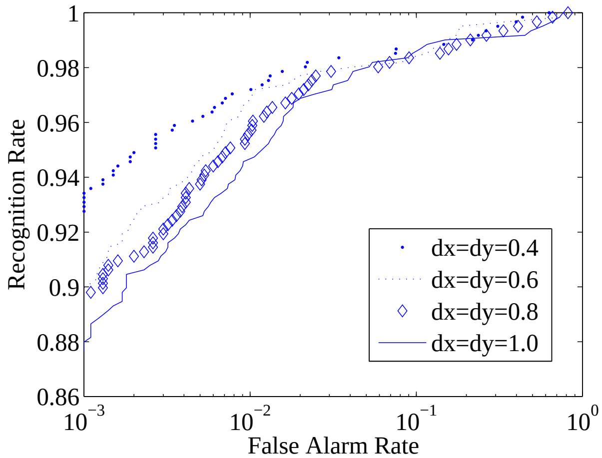

Fig. 3. ROC performance for using different unit difference on the x and y coordinates.

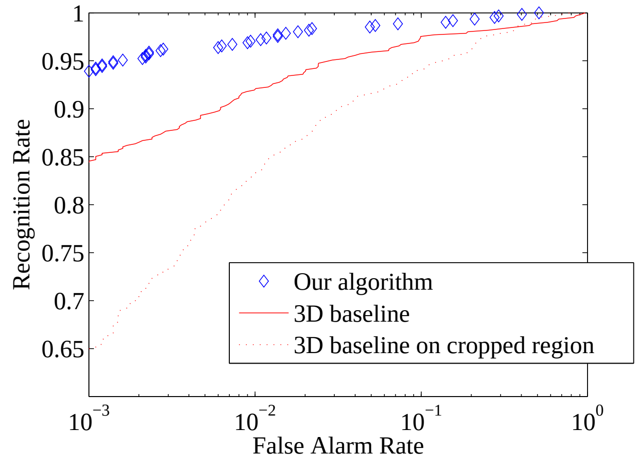

Fig. 4. Comparison with FRGC 3D baseline algorithm. We apply the FRGC 3D baseline algorithm on the normalized depth map and the cropped region shown in Fig 1. Their corresponding ROC curves are shown by the solid line and the dash line respectively.

5. CONCLUSION

The value of summation invariants in the context of 3D face recognition is evaluated in this paper. We compute summation invariants from the nose region of a face as it has prominent shape changes. Compared with the FRGC 3D baseline algorithm, our algorithm obtains higher recognition performance. Furthermore, such good performance can be achieved using only the nose portion of the whole face. In general, the results support the conclusion that summation invariants extract useful information for recognition purposes. The combination of different summation invariants is expected to yield higher recognition performance. We are currently in the process of testing different decision fusion rules.

6. REFERENCES

[1] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches to three-dimensional face recognition," in *Proceedings of ICPR*, 2004, vol. 1, pp. 358–61.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[3] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Transactions on PAMI*, vol. 22, no. 10, pp. 1090–104, 2000.

[4] P. J. Phillips, P. Grother, R. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002," in *International Workshop on AMFG*, 2003, p. 44.

[5] M. Fels and P. J. Olver, "Moving coframes: I. a practical algorithm," *Acta Applicandae Mathematicae*, vol. 51, no. 2, pp. 161 – 213, 1998.

[6] M. Fels and P. J. Olver, "Moving coframes: Ii. regularization and theoretical foundations," *Acta Applicandae Mathematicae*, vol. 55, no. 2, pp. 127 – 208, 1999.

[7] L. Van Gool, P. Kempenaers, and A. Oosterlinck, "Recognition and semi-differential invariants," in *Proceedings of CVPR*, 1991, pp. 454 – 60.

[8] T. Moons, E. J. Pauwels, L. J. Van Gool, and A. Oosterlinck, "Foundations of semi-differential invariants," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 25–47, 1995.

[9] C. E. Hann and M. S. Hickman, "Projective curvature and integral invariants," *Acta Applicandae Mathematicae*, vol. 74, no. 2, pp. 177–193, 2002.

[10] J. Sato and R. Cipolla, "Affine integral invariants and matching of curves," in *Proceedings of 13th ICPR*, Vienna, Austria, 1996, vol. 1, pp. 915–19.

[11] W. Y. Lin, N. Boston, and Y. H. Hu, "Summation invariant and its application to shape recognition," in *Proceedings of ICASSP*, 2005, vol. V, pp. 205–208.

[12] 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," in *Proceedings of CVPR*, 2005, vol. 1, pp. 947–54.