3D Facial Expression Recognition Based on Primitive Surface Feature Distribution

Jun Wang, Lijun Yin, Xiaozhou Wei and Yi Sun
Department of Computer Science
State University of New York at Binghamton, NY, 13902, USA

Abstract

The creation of facial range models by 3D imaging systems has led to extensive work on 3D face recognition [19]. However, little work has been done to study the usefulness of such data for recognizing and understanding facial expressions. Psychological research shows that the shape of a human face, a highly mobile facial surface, is critical to facial expression perception. In this paper, we investigate the importance and usefulness of 3D facial geometric shapes to represent and recognize facial expressions using 3D facial expression range data. We propose a novel approach to extract primitive 3D facial expression features, and then apply the feature distribution to classify the prototypic facial expressions. In order to validate our proposed approach, we have conducted experiments for person-independent facial expression recognition using our newly created 3D facial expression database. We also demonstrate the advantages of our 3D geometric based approach over 2D texture based approaches in terms of various head poses.

1. Introduction

There is a long history of interest in the problem of recognizing human emotion from facial expressions, as well as extensive studies on face perception over the last three decades [8, 17]. Analyzing the emotional expression of a human face requires a number of preprocessing steps which attempt to detect and locate characteristic facial regions, extract facial expression features, and model facial gestures using anatomic information about the face. Although all these steps are equally important, current research mostly concentrates on the facial expression feature detection and description, which is also the focus of this paper.

Facial expression features are mainly represented by three categories: (1) static versus dynamic (or temporal); (2) global versus local (or analytic); and (3) 2D versus 3D. Most research over the past thirty years has been directed towards static/dynamic, analytic, 2D feature extraction [9], focusing primarily on two types of features: 2D geometric features and appearance features from 2D static images [16, 18] or video sequences [4, 3, 31]. Geometric features are described by a set of facial feature points used to derive facial organs' shapes or expressive regions. Appearance features refer to the features exhibited on the skin (e.g., frowns or wrinkles) [22] or action units [8, 6]. Over the past decade, a number of techniques have been successfully developed for facial expression recognition, including optical flow [28, 10] and Gabor wavelets [16], and FACS-feature based techniques [22, 1, 6, 31]. The excellent review of recent advances in this field can be found in [5, 12, 17, 32]. All recent advances have been based on 2D images or videos, and most were primarily concerned with extracting prototypic expressions from frontal or near-frontal views of a face.

Recently, additional work has been done to improve the performance of facial expression recognition under various imaging conditions (e.g., pose, lighting, etc). Some researchers have successfully explored partial 3D information for facial expression recognition, such as multiple-view based [18] and 3D model-based techniques [2, 13, 27, 30]. These methods are still based on 2D data. Because the facial expression actuated by the facial muscle movement results in the facial skin shape variation, it is ideal to model the facial expressions explicitly in a 3D space.

Due to the limitations in describing facial surface deformation when 3D features are evaluated in 2D space, 2D images with a group of feature units may not accurately reflect complex and authentic facial expressions. More importantly, head pose and posture, which are precious cues in conjunction with facial action, reflect a person's real emotion. Since people rarely express emotions without head motion or posture spontaneity, the assumption of frontal images of faces under good illumination is to be unrealistic. Therefore, there is a high demand to represent and recognize facial expressions in 3D space. In this paper, we address the issue regarding 3D global features on the 3D facial surface in order to mitigate the problems posted by 2D based facial expression analysis.

The 3D surface features, reflecting the facial skin "wave", represent the intrinsic facial surface structure associated with the specific facial expressions. Motivated by this fact, we propose a novel geometric feature based facial expression descriptor in the 3-dimensional Euclidean space. Based on the principal curvature information estimated on the 3D triangle mesh model, we apply a surface labeling approach to classify the 3D primitive surface features into twelve basic categories. In order to classify the specific expression, we partition the face surface into a number of expressive regions, and conduct the statistics of the surface primitive label distribution on each region separately. The statistic histograms of the surface labels of all these regions are combined to construct the specific facial expression feature.

Finally, we validate our approach by the recognition experiments on our newly constructed 3D facial expression database. We conduct the experiments in the following several aspects: the performance investigation of our 3D approach; the comparison study with 2D appearance feature based methods (Gabor-Wavelet method and our Topographic Context method), including the front view case and the angle-view case. The experiments are executed in a person-independent manner, which means that the subject being tested has never appeared in the training set. We have selected four classifiers for the classification experiments. Since the paper focuses on 3D facial expression description, the classification algorithm design is not expatiated on in this paper.

The remainder of this paper is organized as follows: In Section 2, we introduce the surface primitive feature analysis based on the computational geometry method. In Section 3, we describe the facial expression feature extraction from the primitive surface features distribution. Section 4 reports the experimental results of the 3D facial expression recognition, followed by a comparison with the 2D appearance feature based approaches in Section 5. Finally, concluding remarks and discussion are given in Section 6.

2. 3D Primitive Feature Analysis

The surface feature analysis is based on the triangle meshes of faces, which are created by a 3D imaging system [15]. The system captures a three dimensional point cloud and generates a meshed surface that models the face as 3D triangle mesh geometry. The 3D surface data points are less than 300 microns apart, providing the capability of distinguishing between small differences regardless of lighting or orientation during the original scan. Figure 1 shows an example of 3D facial expression range models with six prototypic facial expressions. In the following sub-sections, we will give a detailed description of the primitive facial feature estimation and labeling of the range models.
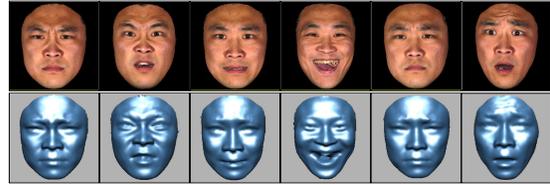

Figure 1. An example of 3D facial range models showing six prototypic expressions, from left to right: *Anger, Disgust, Fear, Happiness, Sadness, and Surprise*. **The textured models are shown in the upper row, and the corresponding shaded models are in the lower row.**

2.1. Principal Curvature Analysis by Local Surface Fitting

The shape information of a surface is "encoded" in its primitive geometric features (such as ridge, ravine, peak, pit, saddle, concave hill, convex hill, etc.), which may be viewed as a digital signature of a facial expression. These features are determined by the surface curvatures and their principal directions. Curvature estimation techniques for triangle meshes could be based on the mesh itself or alternatively on a local smooth approximation. In general, there are broadly two categories: discrete and continuous [11]. The first refers to approximating curvatures by formulating a closed form for differential geometry operators that work directly on the discrete representation of the underlying surface. The latter involves fitting a surface locally, then computing the curvatures by interrogating the fitted surface. Our experiment shows that curvature estimates derived by locally approximating the surface with a smooth polynomial function give better results than the discrete versions.

In order to derive the geometry features of facial surface, we give an analytic form of the regression function for smoothly fitting the triangle mesh surface. Note that although the triangle meshes are defined in a global coordinate system, for easy computation, the fitting procedure is performed in a local coordinate system, which is centered at the vertex to be examined. The local coordinate system for the surface fitting is defined as follows: let p be a vertex on a 3D model $\mathcal{M}$, and $\mathbf{n}_p = (a, b, c)^T$ be the unit normal at vertex p. A set of vertices that are adjacent to p is $\{q_i = (x_i, y_i, z_i)^T\}, i = 1, 2..., m$. A local coordinate system is defined by taking the vertex p as an origin and $(\mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_p)$ as the three axes. $\mathbf{n}_x$ and $\mathbf{n}_y$ are two orthogonal axes, which are arbitrarily selected to form a tangent plane perpendicular to the normal vector $\mathbf{n}_p$ at crossing point p (see Figure 2 for an example). The normal vector can be simply estimated by the mean normal, which is obtained by
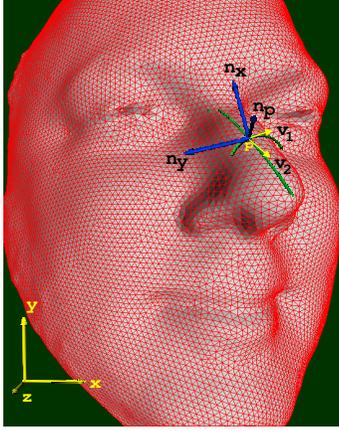
Figure 2. An example of a local coordinate system $(\mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_p)$ defined on a nose surface patch. $\mathbf{v}_1, \mathbf{v}_2$ are the principal directions at point p.

the average of normal values of the polygons sharing the common vertex p.

To fit a smooth polynomial patch onto the local surface and estimate the principal curvatures, we transform the vertices of the local region to the local coordinate system. The resulting vertices $\tilde{\mathbf{q}}_i$ and the corresponding normal vectors $\tilde{\mathbf{n}}_{\tilde{q}_i}$ are expressed as:

$$\begin{aligned} \tilde{\mathbf{q}}_i &= \mathbf{R}^T \cdot (q_i - p) \\ \tilde{\mathbf{n}}_{\tilde{q}_i} &= \mathbf{R}^T \cdot \mathbf{n}_{q_i} \end{aligned} \quad (1)$$

where the rotation matrix is obtained by $\mathbf{R} = [\mathbf{n}_x \ \mathbf{n}_y \ \mathbf{n}_p]$. The vertex p and its normal $\mathbf{n}_p$ are transformed to the origin $\tilde{p} = (0,0,0)^T$ and the unit vector $\tilde{\mathbf{n}}_p = (0,0,1)^T$ along the positive z axis, respectively.

The approximating polynomial surface is at least second order. Considering the fitting accuracy and the computation complexity, we choose a cubic-order approximation method. Similar to the method in [14], we define the fitting function in the form of

$$z(\tilde{x}, \tilde{y}) = \mathbf{U} \cdot \mathbf{X} \quad (2)$$

where the variable $\mathbf{U}$ and the coefficient $\mathbf{X}$ are defined as:

$$\begin{aligned} \mathbf{U} &= (\tfrac{1}{2}\tilde{x}^2 \ \tilde{x}\tilde{y} \ \tfrac{1}{2}\tilde{y}^2 \ \tilde{x}^3 \ \tilde{x}^2\tilde{y} \ \tilde{x}\tilde{y}^2 \ \tilde{y}^3) \\ \mathbf{X} &= (A \ B \ C \ D \ E \ F \ G)^T \end{aligned} \quad (3)$$

Given a set of neighbor vertices $\{\tilde{q}_i = (\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)^T\}$, $(i = 1, 2, ..., m)$ and their corresponding normals $\{\tilde{\mathbf{n}}_{\tilde{q}_i} =$

$(a_i, b_i, c_i)^T\}$, we can establish $3m$ equations to solve the seven parameters of $\mathbf{X}$. When the vertex p has more than three neighbor vertices, the parameters $\mathbf{X}$ can be approximated using the least-square fitting method. With the regressed local fitting function $z(\tilde{x}, \tilde{y})$, the Weingarten matrix for the surface patch becomes:

$$\mathbf{W} = \begin{bmatrix} \frac{\partial^2 z(\tilde{x},\tilde{y})}{\partial \tilde{x}^2} & \frac{\partial^2 z(\tilde{x},\tilde{y})}{\partial \tilde{x} \partial \tilde{y}} \\ \frac{\partial^2 z(\tilde{x},\tilde{y})}{\partial \tilde{x} \partial \tilde{y}} & \frac{\partial^2 z(\tilde{x},\tilde{y})}{\partial \tilde{y}^2} \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix} \quad (4)$$

After the eigenvalue decomposition, the principal directions in the local coordinates $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ can be estimated.

$$\mathbf{W} = (\tilde{\mathbf{v}}_1 \ \tilde{\mathbf{v}}_2) \cdot diag(\lambda_1 \ \lambda_2) \cdot (\tilde{\mathbf{v}}_1 \ \tilde{\mathbf{v}}_2)^T \quad (5)$$

where λ_1 and λ_2 are the eigenvalues and $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ are the orthogonal eigenvectors. If $|\lambda_1| > |\lambda_2|$, $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ are in the directions with the maximum curvature and the minimum curvature, respectively. Figure 3(a)(b) shows an example of the principal direction estimation on a facial range model. Since the principal directions are represented in the local coordinate system, to obtain a global view of the principal directions, $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ must be rotated back to the global coordinate system, as formulated by

$$\mathbf{v}_1 = \mathbf{R} \cdot \tilde{\mathbf{v}}_1 \ , \ \mathbf{v}_2 = \mathbf{R} \cdot \tilde{\mathbf{v}}_2 \quad (6)$$

It is worth noting that our surface labeling criteria rely on the surface principal curvatures, the principal directions as well as the surface gradient. After the transformation from the local coordinate system to the original global coordinate system, we can derive the gradient using the normal direction of each surface point. Let $z(x,y)$ be the fitting function of a surface patch centered at p, the normal direction $\mathbf{n}_p = (a,b,c)^T$ can be written in the form of $(-\frac{\partial z(x,y)}{\partial x}, -\frac{\partial z(x,y)}{\partial y}, -1)^T$. The gradient magnitude $\|\nabla z(x,y)\|$ at p is then calculated as:

$$\|\nabla z(x,y)\| = \sqrt{\left[\frac{\partial z(x,y)}{\partial x}\right]^2 + \left[\frac{\partial z(x,y)}{\partial x}\right]^2} = \sqrt{\left[-\frac{a}{c}\right]^2 + \left[-\frac{b}{c}\right]^2} \quad (7)$$

The principal curvature analysis produced a set of attributes $\{\|\nabla z\|, \mathbf{v}_1, \mathbf{v}_2, \lambda_1, \lambda_2\}$, which describes the surface property at each vertex. Every vertex can be classified according to a primitive feature classification rule, which will be explained in the next sub-section.

2.2. Primitive 3D Surface Feature Labeling

The principal curvatures λ_1, λ_2 represent the maximum and the minimum degrees of bending of a surface, $\mathbf{v}_1$ and $\mathbf{v}_2$ indicate the surface principal directions, and $\|\nabla z\|$ reflects the steepness of the surface. Using these geometric attributes, we are able to classify every vertex into one of the primitive categories. In other words, we can symbolize
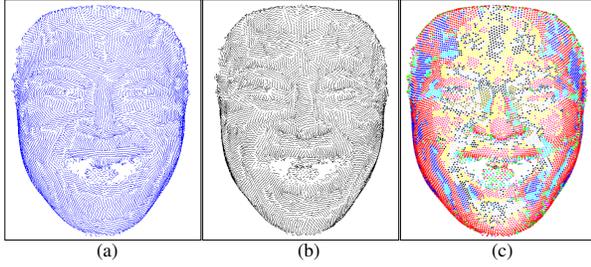
Figure 3. From left to right: minimum curvature direction map, maximum curvature direction map and primitive label map.

the geometric surface by classifying and labeling the primitive features. This symbolization process can be realized by mapping the primitive surface features from a 3D geometric space to a discrete label space, which is formulated as:

$$\mathcal{Q} : \{\|\nabla z\|, \mathbf{v}_1, \mathbf{v}_2, \lambda_1, \lambda_2\} \Longrightarrow \{L_m\}, m = 1, 2, ..., M \tag{8}$$

where $\mathcal{Q}$ is a pre-defined classification rule and $\{L_m\}$ is a set of surface categories with a total of M types to be identified.

There are several existing rules for primitive surface feature classification. For example, (1) the shape index based classification [7]; (2) Tanaka's method [21], which categorized eight distinct features to describe local shapes according to the sign of two principal curvatures; (3) the topographic classification method [23, 26], which has been used for analyzing the 3D topographic surfaces of gray level images. Depending on the classification fineness, the maximal twelve distinct primitive features can be defined. They are peak, pit, flat, ravine, ridge, saddle (including ridge saddle and ravine saddle) and hill (including convex hill, concave hill, concave saddle hill, convex saddle hill and slope hill).

In order to scrutinize the facial expression surface details, we use the twelve distinct primitive surface features to represent the facial expressions. Similar to the classification rule used in [23], we extend the method to the application for the real 3D facial surface labeling. The labeling process is based on the feature values of $\{\|\nabla z\|, \mathbf{v}_1, \mathbf{v}_2, \lambda_1, \lambda_2\}$ in the global coordinate system. To do so, two thresholds, T_G and T_λ, are defined. They are used to evaluate whether the gradient magnitude and the principal curvatures are trivial enough to be ignored as zero. The thresholds are calculated based on the mixed error criteria [26]:

$$T_G = \max[\epsilon, \epsilon \cdot z(x, y)]$$
$$T_\lambda = \max[\epsilon, \epsilon \cdot \|W\|_\infty] \tag{9}$$

where $\|W\|_\infty = \max[|A| + |B|, |B| + |C|]$ (see Equation

4) and ϵ is a user-specified parameter. In our method, we use $\epsilon = 0.001 \cdot s$, where s is the average distance from each vertex to the center of the individual 3D facial model.

The classification rule for twelve primitive surface labels is expounded in Table 1. In general, if $\|\nabla z\| < T_G$ or there is a zero crossing in the direction of the maximum curvature, one of the non-hillside labels is assigned; otherwise, one of the hill-side labels is assigned using the rule defined in the table. Figure 3(c) shows an example of the labeling result from a 3D facial expression range model.

λ_1	λ_2	Hillside Label	Non-Hillside Label
$|\lambda_1| < T_\lambda$	$|\lambda_2| < T_\lambda$	flat	slope hill
$\lambda_1 < -T_\lambda$	$\lambda_2 < -T_\lambda$	peak	convex hill
$\lambda_1 < -T_\lambda$	$|\lambda_2| < T_\lambda$	ridge	convex hill
$\lambda_1 < -T_\lambda$	$\lambda_2 > T_\lambda$	ridge saddle	convex saddle hill
$\lambda_1 > T_\lambda$	$\lambda_2 < -T_\lambda$	ravine saddle	concave hill
$\lambda_1 > T_\lambda$	$|\lambda_2| < T_\lambda$	ravine	convex hill
$\lambda_1 > T_\lambda$	$\lambda_2 > T_\lambda$	pit	
$\lambda_1 > T_\lambda$	$\lambda_2 < -T_\lambda$		concave saddle hill

Table 1. Classification rule of primitive 3D surface labels.

3. 3D Expression Description Based on Primitive Label Distribution

After the labeling process, the facial expressions can be described by the distribution of the labels over the local or entire facial region. Intuitively, every facial expression is a result of facial muscle actuation, reflecting the facial surface variation. Such a variation results in the different distributions of primitive surface labels. This fact suggests that the primitive label distribution could directly link to a distinct facial expression. In other words, the same type of facial expression is expected to share the similar primitive label distribution with a certain robustness, given a sufficient resolution of the range model.

To find an explicit representation of the fundamental structure of facial surface details, we investigate the statistical distributions of the primitive surface labels in seven expressive facial regions, which are defined according to the neuro-anatomy knowledge of configuration of facial muscles and their dynamics [20]. As shown in Figure 4, sixty-four fiducial points are defined on the facial surface, and accordingly, the seven expressive local regions are constructed based on these key points. Note that the interiors of mouth and eyes are currently not included in the seven local regions. The reasons are twofold: (1) the interiors of mouth and eyes are isolated from the facial skin. They can be
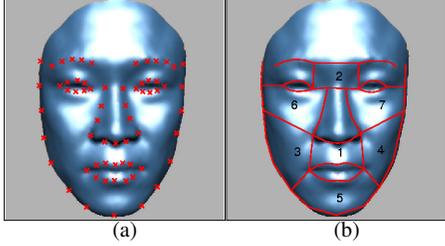
Figure 4. 64 **facial fiducial points and** 7 **selected facial regions.**

treated as three separated objects. These "holes" are complementary areas of the facial skin area. The change of their shapes is directly reflected on the change of the surrounding skins which have been included in the seven expressive regions; (2) although tongues and eye-balls can be viewed as the signatures of facial expressions, their appearances may not be as stable as the facial skin due to the *dark hole effect* existing in the current 3D imaging systems.

In short, the selected seven local regions cover the most expressive areas on a human face. In each selected region, we identify the expression signature by calculating the histogram distribution of primitive labels. Such a distribution is described as follows:

$$\mathbf{r}_i = \left[\frac{n_{i1}}{n_i}, \cdots, \frac{n_{im}}{n_i}, \cdots, \frac{n_{iM}}{n_i} \right] \qquad (10)$$

where n_{im} is the number of vertices which are labeled by the label type L_m, and n_i is the total number of vertices in the ith local region ($n_i = \sum_{m=1}^{M} n_{im}$). $M = 12$ is the number of the primitive label categories.

The combination of seven histogram distributions of the selected entire regions generates a unique expression descriptor for a specific expression. As a result, the expression is described by the primitive surface feature distribution (PSFD), which is expressed by

$$\mathbf{E} = [\mathbf{r}_1, \cdots, \mathbf{r}_i, \cdots, \mathbf{r}_K] \qquad (11)$$

where K is the number of expressive regions ($K{=}7$ in our experiment).

4. Recognition Experiments

4.1. Database

We constructed a 3D facial expression database for our experiment. The 3D range data is scanned by a 3DMD static digitizer [15], which uses a random light pattern projection

in the speckle projection flash environment . The model resolution is in the range of 20,000 polygons to 35,000 polygons, depending on the size of the face being scanned. The facial expressions with six universal emotional states, *Anger, Disgust, Fear, Happiness, Sadness* and *Surprise*, are sampled in four different levels of intensity (e.g., from less pronounced to more pronounced). In our experiment, we used the data captured from 60 subjects with two high-intensity models for each expression. The test is based on the six prototypic expressions.

Each range facial model consists of a meshed surface model and an associated texture image. The head pose can be estimated by a triangle plane determined by the inner corners of two eyes and a nose tip. Given the head pose, the original scan is processed by rotating the model to the frontal view. This pre-processing results in a *working model*, on which we manually labeled 64 fidicual vertices for the expression recognition experiments. Table 2 gives a summary of the data set that we used for our experiments. A detail description of the database construction, post-processing, and organization can be found in [29].

# of subjects	# of expression types	# of samples for each subject of each expression	total # of samples
60	6	2	720

Table 2. Summary of the 3D **facial expression data set used in our experiment.**

4.2. Recognition Results

Our facial expression recognition experiments are carried out in a person-independent manner, which is believed to be more challenging than a person-dependent approach [17]. We randomly partitioned the 60 subjects into two subsets: one with 54 subjects for training and the other with 6 subjects for test. The experimental paradigm guarantees that any subject used for testing does not appear in the training set because the random partition is based on the subjects rather than the individual expression. Four popular classifiers: Quadratic Discriminant Classifier (QDC), Linear Discriminant Analysis (LDA), Naive Bayesian Classifier (NBC), and Support Vector Classifier (SVC) with RBF kernel are used in the experiments. The tests are executed 20 times on each classifier with different partitions to achieve a stable generalization recognition rate. The entire process guarantees that every subject is tested at least once for each classifier. For each round of the test, all the classifiers are reset and re-trained from the initial state.

Classifier	QDC	LDA	NBC	SVC
Recognition rate	74.5%	83.6%	71.7%	77.8%

Table 3. Results of person-independent expression classification using the 3D-PSFD method.

Table 3 shows the average correct recognition rates of the four classifiers. The LDA classifier achieves the highest correct recognition rate with an accuracy of 83.6%. The confusion matrix of the average case for the LDA classifier is shown in Table 4. The expressions of *Happiness* and *Surprise* are well identified with accuracies of 95.0% and 90.8%, respectively. *Anger*, *Sadness*, *Fear* and *Disgust* have comparatively lower recognition rates. The misclassification rate between *Anger* and *Sadness* is around 19.6% (= 8.3%+11.3%), while that between *Fear* with *Happiness* is around 16.3%.

Input\Output	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	80.0%	1.7%	6.3%	0.0%	11.3%	0.8%
Disgust	4.6%	80.4%	4.2%	3.8%	6.7%	0.4%
Fear	0.0%	2.5%	75.0%	12.5%	7.9%	2.1%
Happiness	0.0%	0.8%	3.8%	95.0%	0.4%	0.0%
Sadness	8.3%	2.5%	2.9%	0.0%	80.4%	5.8%
Surprise	1.7%	0.8%	1.2%	0.0%	5.4%	90.8%

Table 4. Confusion Matrix of the average case of LDA classifier for person-independent expression recognition.

5. Comparison Study

In this section, we compare the proposed 3D primitive feature distribution method (3D-PSFD) with two 2D appearance feature based methods. One is the well-known Gabor-wavelet (GW) approach [16] and the other is our recently developed Topographic Context (TC) approach [24].

(1) In the Gabor-wavelet (GW) based approach, a set of multi-scale and multi-orientation coefficients are calculated to describe the appearance variations in the facial region. We applied 6×3 complex Gabor-wavelet (6 orientations and 3 spatial resolutions) on 34 fiducial points. The coefficients of the real and imaginary parts can be computed as:

$$G_+(x,y;\omega,\theta) = \frac{\omega^2}{\sigma^2} \cdot e^{-\frac{\omega^2(x^2+y^2)}{2\sigma^2}} \cdot \{\cos[\omega(x\cos\theta+y\sin\theta)] - e^{-\frac{\sigma^2}{2}}\}$$
$$G_-(x,y;\omega,\theta) = \frac{\omega^2}{\sigma^2} \cdot e^{-\frac{\omega^2(x^2+y^2)}{2\sigma^2}} \cdot \sin[\omega(x\cos\theta+y\sin\theta)] \quad (12)$$

where we let $\sigma = \pi$. The three selected spatial scales are $\{\pi/4, \pi/8, \pi/16\}$ and the 6 orientations selected are $\{0, \pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6\}$. The feature extracted at pixel (x, y) by a certain Gabor-wavelet kernel with parameters $\{\omega, \theta\}$ is the amplitude of the real and imaginary coefficients $G = \sqrt{G_+^2 + G_-^2}$. Therefore in total, we extracted $18 \times 34 = 612$ wavelet features for each image.

(2) In our existing work, we developed a Topographic Context (TC) based approach for facial expression recognition based on 2D static images. The TC expression features are the topographic primal sketch features inherent in the 2D facial images. We use such appearance features to describe the distinct facial expressions. Similar to our 3D based approach, we used the feature statistics to represent the 2D facial expressions. It has proved to be robust to facial landmark detection as a result of its intrinsic statistics property (see details in [24]).

The comparison study with the above two methods is conducted under various head pose conditions, including two cases: frontal view and non-frontal views.

5.1. Case 1: Frontal View

Our frontal-view images are generated from the texture-mapped working models. All these images are normalized to 256×256 pixels. Several examples are shown in the top row of Figure 1. There are 60 subjects with a total of 720 frontal-view facial images used for the test. The experiments are executed in a person-independent manner similar to the strategy used in Section 4.2.

Table 5 reports the correct recognition rates using these two methods. We found the Gabor-wavelet approach performs poorly when using the SVC classifier, which is not comparable to the results from the other three classifiers.

Method \ Classifier	QDC	LDA	NBC
Topographic context method	73.8%	79.2%	70.9%
Gabor-wavelets method	72.3%	74.1%	62.1%

Table 5. Results of person-independent expression classification using GW and TC.

Comparing to the performance shown in Table 3, the 3D-PSFD method is superior to the 2D appearance feature based methods when classifying the six prototypic facial expressions.

5.2. Case 2: Non-frontal View

In this section, we compare the performance of our 3D geometric based approach (PSFD) with the 2D appearance

feature based approaches under different head pose conditions. We obtained the head pose information from the original face scans by estimating the triangle face-plane formed by the inner corners of two eyes and the tip of nose. Knowing the 3D head pose, we are able to rotate the 3D model to a frontal view position or arbitrary pose positions without losing any geometric feature information. However, if viewed in the 2D projection plane, the appearance of 2D facial images varies dramatically when the head pose is arbitrarily changed. The difficulty to recover the missing appearance information from 2D images makes 2D based facial expression recognition sensitive to head pose variations.

Using our face range models, we have generated facial expression images under different views, corresponding to different head poses. This is done by rotating the models to a certain degree and generating the face images in that view by texture-mapping. We generated the face images under views of $\pm 10°$, $\pm 20°$, $\pm 30°$, and $\pm 40°$ for each orientation (pitch and yaw rotation), resulting in 720 images for each view. Figure 5 shows the examples of different facial appearances under different views.
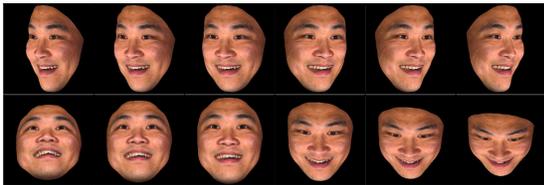

Figure 5. Facial expression images with different head rotation. Top row is yaw rotation and bottom row is pitch rotation. From left to right, the rotation angle is $-40°$**,** $-30°$**,** $-20°$**,** $20°$**,** $30°$ **and** $40°$**.**

We choose to use the LDA classifier to evaluate the expression recognition performance under different head poses because in most cases LDA outputs the best recognition result. For all three algorithms (PSFD, GW, and TC), the LDA classifier is trained using the front view face data.

The recognition results with respect to different poses (e.g., pitch and yaw rotations) are shown in Figure 6. From the figure, we can see that the average recognition rates of the 2D appearance feature based methods degrade considerably as head rotation increases, especially the Gabor-wavelet method. The extraction of Gabor-wavelet coefficients is based on the selected fiducial points. The pose change alters the distribution of these fiducial points, and results in significant distortion of Gabor-wavelet features. On the contrary, the 3D-PSFD based approach makes the 3D primitive feature distribution invariant to the pose variations because the 3D geometric features are view-independent.
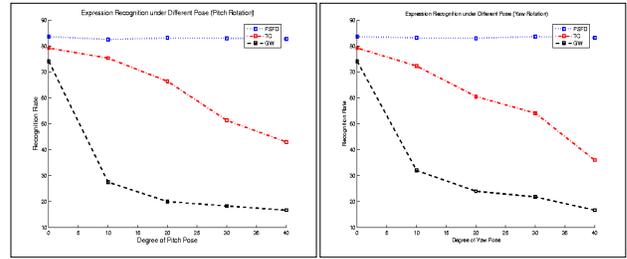

Figure 6. Comparison of the recognition performance under different head orientations. left: pitch; right: yaw.

6. Concluding Remarks

In this paper, we investigated the issue of 3D facial expression representation and recognition. To the best of our knowledge, this is the first attempt to recognize facial expressions using range data in a complete 3D space. We have proposed to extract and label the primitive 3D surface features, and derive their statistical distributions to represent the distinct prototypic facial expressions. We used the primitive surface feature distribution (PSFD) as the signature to distinguish facial expressions, and conducted expression recognition experiments using the person-independent strategy. Compared to the existing 2D static image based approaches (e.g., GW and TC methods), our 3D range data based approach shows superior performance as a result of the lighting and orientation invariance of 3D geometric features. The experiments show encouraging results in recognizing six prototypic facial expressions under various head-pose conditions.

There are some limitations in the current work:

(1) Our current data set contains only static expression models. A database including dynamic 3D facial expression sequences is needed in order to study the subtle skin movement associated with facial expressions in 3D space. With the emergence of dynamic 3D imaging systems [25], it is possible to investigate the feasibility of tracking action units in 3D space to further enhance the current FACS based facial expression recognition technique. Because the discriminability of 3D facial expressions is dependent on the resolution of the 3D mesh, higher resolution models are required in order to improve the recognition performance.

(2) The current work involves the pre-processing of range data which requires manual selection of the surface fiducial points. In order to realize an automatic system for 3D facial expression analysis, algorithms for automatic detection of 3D surface features must be developed. The existing approaches (e.g., free-form based [25], etc.) are promis-

ing for this purpose.

The above issues give rise to our future research directions in order to improve facial expression analysis in the 3D space. In addition, we will investigate integrating 3D geometric shape and 2D texture information to improve our current approach.

7. Acknowledgement

References

[1] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I.Fasel, and J. Movellan. Recognizing facial expressions: machine learning and application to spontaneous behavior. In *IEEE CVPR2005*, San Diego, CA. 2005.

[2] B. Braathen, M. Bartlett, G. Littlewort, E. Smith, and J. Movellan. An approach to automatic recognition of spontaneous facial actions. In *Proc. of Int. Conf. on FGR*, 2002.

[3] Y. Chang, C. Hu, and M. Turk. Probabilistic expression analysis on manifolds. In *CVPR*, Washington DC, USA, 2004.

[4] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang. Facial expression recognition from video sequences: temporal and static modeling. *CVIU*, 91:160–187, 2003.

[5] R. Cowie, , E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor. Emotion recognition in human computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, 2001.

[6] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. *IEEE Trans. on PAMI*, 21:974–989, 1999.

[7] C. Dorai and A. Jain. Cosmos - a representation scheme for 3d free-form objects. *IEEE Trans. on PAMI*, 19(10):1115–1130, 1997.

[8] P. Ekman and W. Friesen, editors. *The facial action coding system: a technique for the measurement of facial movement*. Consulting Psychologists Press, San Francisco, 1978.

[9] P. Ekman, T. Huang, T. Sejnowski, and J. Hager. *Final report to NSF of the planning workshop on facial expression understanding*. Human Interaction Lab., UC at San Francisco, 1993.

[10] I. Essa and A. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. on PAMI*, 19:757–763, 1997.

[11] G. Farin. *Curves and Surfaces for Computer Aided Geometric Design*. 5th ed., Morgan-Kaufmann, 2001.

[12] B. Fasel and J. Luttin. Automatic facial expression analysis: Survey. *Pattern Recognition*, 36:259–275, 2003.

[13] S. Gokturk, J. Bouguet, C. Tomasi, and B. Girod. Model-based face tracking for view-independent facial expression recognition. In *Proc. of Int. Conf. on FGR*, 2002.

[14] J. Goldfeather and V. Interrante. A novel cubic-order algorithm for approximating principal direction vectors. *ACM Trans. on Graphics*, 23:45–63, 2004.

[15] Inc.3dMd. *http://www.3q.com*. 2005.

[16] M. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Trans. on PAMI*, 21:1357–1362, 1999.

[17] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Trans. on PAMI*, 22:1424–1445, 2000.

[18] M. Pantic and L. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Trans. on SMC-Part B: Cybernetics*, 34:1449–1461, 2004.

[19] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Conf. on CVPR, San Diego, CA*, 2005.

[20] W. Rinn. The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, 95:52–77, 1984.

[21] H. Tanaka, M. Ikeda, and H. Chiaki. Curvature-based face surface recognition using spherical correlation. In *IEEE Conf. on FGR*, pages 372–377, 1998.

[22] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE Trans. on PAMI*, 23:1–9, 2001.

[23] O. Trier, T. Taxt, and A. Jain. Data capture from maps based on gray scale topographic analysis. In *The Third International Conference on Document Analysis and Recognition*, Montreal, Canada, 1995.

[24] J. Wang and L. Yin. Facial expression representation and recognition from static images using topographic context. In *Technical Report, Department of Computer Science, SUNY at Binghamton*, Nov., 2005.

[25] Y. Wang, X. Huang, C. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, and P. Huang. High resolution acquisition, learning and transfer of dynamic 3d facial expressions. In *EUROGRAPHICS 2004*, 2004.

[26] L. T. Watson, T. J. Laffey, and R. M. Haralick. Topographic classification of digital image intensity surfaces using generalized splines and the discrete cosine transform. *Computer Vision, Graphics and Image Processing*, 29:143–167, 1985.

[27] Z. Wen and T. Huang. Capturing subtle facial motions in 3d face tracking. In *ICCV*, 2003.

[28] Y. Yacoob and L. Davis. Recognizing human facial expression from long image sequences using optical flow. *IEEE Trans. on PAMI*, 16:636–642, 1996.

[29] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato. A 3d facial expression database for facial behavior research. In *7th Int. Conf. on FGR*, Southampton, UK, 2006.

[30] L. Zalewski and S. Gong. Synthesis and recognition of facial expressions in virtual 3d views. In *IEEE 6th Inter. Conf. on FGR*, 2004.

[31] Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Trans. on PAMI*, 27(5):699–714, May 2005.

[32] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), Dec. 2003.