# Quantifying Facial Expression Abnormality in Schizophrenia by Combining 2D and 3D Features

Peng Wang [1], Christian Kohler [2], Fred Barrett [2], Raquel Gur [2], Ruben Gur [2], Ragini Verma [1]
[1] Department of Radiology
University of Pennsylvania
Philadelphia, PA 19104
{wpeng@ieee.org, ragini.verma@uphs.upenn.edu }

[2] Department of Psychiatry
University of Pennsylvania
Philadelphia, PA 19104
{kohler, fbarrett, raquel, gur}@bbl.med.upenn.edu

## Abstract

*Most of current computer-based facial expression analysis methods focus on the recognition of perfectly posed expressions, and hence are incapable of handling the individuals with expression impairments. In particular, patients with schizophrenia usually have impaired expressions in the form of "flat" or "inappropriate" affects, which make the quantification of their facial expressions a challenging problem. This paper presents methods to quantify the group differences between patients with schizophrenia and healthy controls, by extracting specialized features and analyzing group differences on a feature manifold. The features include 2D and 3D geometric features, and the moment invariants combining both 3D geometry and 2D textures. Facial expression recognition experiments on actors demonstrate that our combined features can better characterize facial expressions than either 2D geometric or texture features. The features are then embedded into an ISOMAP manifold to quantify the group differences between controls and patients. Experiments show that our results are strongly supported by the human rating results and clinical findings, thus providing a framework that is able to quantify the abnormality in patients with schizophrenia.*

## 1. Introduction

Facial expressions have been widely used in clinical research to study the affective and cognitive states, and psychopathology of an individual. Specifically, the facial expression analysis has played a major role in the study of schizophrenia, which is a neuropsychiatric disorder characterized by deficits in emotional expressiveness [25]. Patients with schizophrenia are known to have impaired performance in emotion processing, both in terms of recognizing and expressing emotions. Patients with schizophrenia often demonstrate either or both types of impairment in facial expressions: "flat affect" (a severe reduction in emotional expressiveness [14]) and "inappropriate affect" (inappropriate expression to intended emotions). In this paper, we quantify the facial expression abnormality of patients with schizophrenia by applying computer vision and pattern recognition techniques and using both 3D surface data and 2D texture images.

Clinicians currently rely on manual and subjective methods to rate expressions, and the clinical research in schizophrenia has focused on the perception and recognition capabilities of the patients, and not so much on the way in which patients express emotions differently from healthy controls. This has led to the need for sophisticated Automated Facial Expression Analysis (AFEA) methods, which allow computers to analyze facial expressions [23, 11, 29]. The merits of automated facial expression analysis are twofold: AFEA can reduce the intensive human intervention that is usually subjective, time-consuming and error prone; AFEA can also provide quantitative analysis results. However, most of current AFEA methods mainly focus on recognizing extreme and posed expressions, which are not often seen in the real world. Such methods would have difficulties in analyzing facial expressions of patients with schizophrenia because of the expression impairments. Example images are shown in Figure 1 (a) and (b), where the patient failed to express when being asked to express sadness and anger. Figure 1 (c) and (d) show other cases where the expressions are not appropriately associated with the intended emotions. Overall, the two deficits underline the fact that the expression changes in patients with schizophrenia are usually very subtle and do not follow the normal pattern of expressions, thus being difficult to recognize by existing facial expression recognition methods.

This paper presents a framework that addresses the challenging problem of quantifying the group differences between facial expressions of healthy controls and patients with schizophrenia. Our method includes two main parts.

1-4244-1180-7/07/$25.00 ©2007 IEEE

First, features are extracted from 3D surface data and 2D images to capture facial expression changes. These features include 2D geometric features, 3D curvature features, and moment invariants combining both 3D geometry and 2D textures. Second, the extracted features are embedded into an ISOMAP manifold, in which the distances of data to the control manifold are measured to quantify the group difference between healthy controls and patients.

Our combined features outperform either 2D geometric or texture features in better recognizing actors' expressions, which demonstrates that the features can capture facial changes well. Also, the quantification in feature manifold is able to identify subtle group differences between patients and controls expression-wisely. The quantitative results obtained from our method are consistent with the human rating results and clinical findings. Furthermore, a patient/control database for the clinical study of schizophrenia has been collected in this research.

(a) (b) (c) (d)

Figure 1. Images of patients with schizophrenia when being asked to express: (a) sadness. (b) anger. (c) anger. (d) disgust

The rest of the paper is organized as follows. In Section 2, the previous work is reviewed. We then introduce the protocol of data acquisition in Section 3. Our feature extraction method is introduced in Section 4. We present the methodology and experimental results of analyzing facial expression abnormality in schizophrenia in Section 5. We conclude the paper in Section 6.

## 2. Previous Work

### 2.1. Clinical Facial Expression Analysis

In clinical research, facial expressions are usually described in two ways: expressions as combinations of action units and universal expressions. The Facial Action Coding System (FACS) has been developed to describe facial expressions using a combination of action units (AU) [8]. Each action unit corresponds to specific muscular activity that produces momentary changes in facial appearance. The universal expression is studied as a whole representation of a specific type of internal emotion, without breaking up expressions into muscular units. Most commonly studied universal expressions include happiness, anger, sadness, fear, and disgust. In this study, universal expressions are analyzed using the Facial Expression Coding System (FACES), which has been designed to code facial behaviors with multiple measurements, and to measure the duration and valence of universal expressions [19]. Clinical studies of schizophrenia analyze emotions usually use a combination of these methods. In this study, we compare the results of our computerized method with those of rating from FACES.

### 2.2. Automated Facial Expression Analysis

Many automated facial expression analysis methods have been developed [23, 11], while most of them focus on the expression recognition. These methods can be categorized based on the data and features they use, or the classifiers created for the expression recognition. In summary, the classifiers include Nearest Neighbor classifier [10], Neural Networks [30], SVM [2], Bayesian Networks [4, 31], AdaBoost classifier [2], Hidden Markov Model [20], and rule based classifier [24]. The data used for AFEA can be 2D images, video data, and 3D surface data, and there are different feature extraction methods for each type of data. We now review some typical features as follows.

The commonly used 2D features include geometric features and texture features. Geometric features represent the spatial changes caused by facial expressions. Tian et al. group the geometric features into permanent and transient features [30]. The permanent features include position of lips, eyes brow, cheek, and any furrows that have become permanent with age. The transient features include facial lines and furrows that are not present at rest but appear with facial expressions. The texture features include image intensity[1], image difference [10], edges [30], Gabor wavelets [22], and manifold features [3]. Both PCA features and image difference require precise alignment between images, which is difficult in real applications. The edge features are also difficult to detect for subtle expression changes. In [22], approximately 1,000 Gabor wavelet coefficients are extracted from a face grid for facial expression recognition as well as for race and sex recognition. In [2], the Gabor wavelets are selected at different locations and orientations using an AdaBoost algorithm, and are then used to train a SVM for each facial expression or action unit. Furthermore, the experiments in [38, 1] demonstrate that the fusion of appearance features (Gabor wavelet or PCA features) and geometric features can provide better accuracy than only using either of them.

In video-based methods, the motion information is often used to describe the facial deformations caused by expressions. The motion information can be calculated from optical flow methods [35, 36], or deformable models [9, 17]. In the work of Yacoob et al. [35], each facial expression sequence is divided into three segments: the beginning, the apex and the ending. Rules are defined to determine the temporal model of facial expressions. Yeasin et al. present

a two-stage method for recognizing facial expressions in video [36]. Optical flow is first computed to recognize facial expressions at each frame, then the output at frame level is fed to a discrete HMM for expression recognition in the level of video. Besides motion information, the manifold features are used for expression recognition in videos by applying a sampling-based probabilistic tracking scheme [3].

Recently, the use of 3D surface data for facial expressions analysis has received attention as the 3D data provides fine geometric information that is invariant to pose and illumination changes. Yin et al. present a method to recognize facial expression using primitive surface features [34]. The method labels the surface data into twelve primitive features based on the geometry. The label distributions on facial regions are then used to recognize expressions. Their results show that 3D surface data can improve the recognition accuracy over only using texture features. However, as several other methods, their method performs best on the posed expressions of high intensity. Indeed, all the methods described above suffer from this drawback and hence cannot be used for recognizing expression changes in patients, which are of low intensity and may not follow the regular pattern of facial changes as in healthy people. Classification based methods are unable to perform well on such data as they have been trained on very different data sets. In the following sections, we provide a framework to quantify the facial expression abnormality of patients with schizophrenia, by combing 2D and 3D features.

## 3. Data Acquisition

There are some existing facial expression databases, including the Japanese Female Facial Expression Database [22], the CMU AU-Coded Facial Expression Database [16], the Rutgers/UCSD Facial Action Coding System database [2], and the Binghamton University 3D Facial Expression (BU-3DFE) Database [37]. All the databases are mainly designed for automated facial expression recognition. None of them are designed for the clinical studies, especially, the study of neuropsychiatric disorders such as schizophrenia, as they mainly comprise of posed expressions which actually do not follow the true trend of expression changes, and usually contain expressions of high intensity.

In our study, a database consisting of facial expressions of both patients with schizophrenia and healthy controls has been acquired under the supervision of psychiatrists. All the participants, including patients and controls, are chosen in pairs matched for the age, race, and gender. The pairings will be used later in pairwise statistical analysis to ensure that statistical effects are only due to the differences between patients and controls, and are not biased by age, race, or gender. Each participant goes through two types of sessions: posed and evoked expressions. In the posed session, participants are asked to express five types of emotions, including happiness, anger, fear, sadness and disgust, at mild, moderate, and peak levels, respectively. In the evoked session, participants are individually guided through vignettes which are provided by the participants themselves and describe a situation in their life pertaining to each emotion. In order to elicit evoked expressions, the vignettes are recounted back to the participants by the psychiatrist, who guides them through all the three levels of expression intensity for each emotion. Recruiting participants in this study requires a lot of efforts because it needs to strictly follow protocols, and it is difficult to find paired participants. The patient/control database currently contains 24 participants, and will include more participants in the future.

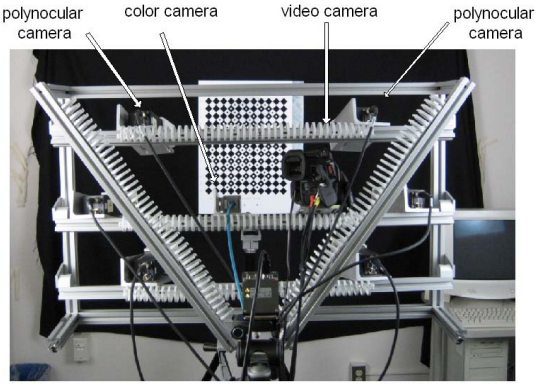



Figure 2. The capturing system setting

Images and videos are acquired during the interview with participants. The capturing setting is illustrated in Figure 2. Participants sit on a well-adjusted chair, facing the cameras. There are six polynocular stereo cameras and one color camera, which are calibrated to provide texture and geometric features [13]. Figure 3 shows examples of 3D surface data and a texture image mapped on the surface. The reconstruction accuracy is further improved by projecting a random IR pattern on the polynocular images and using fish-scales methods [26]. The surface reconstruction errors are measured by comparing reconstruction results with laser scanning results. Experiments show that about $90\%$ of surface points have error less than 2.5 mm, which is accurate enough to measure expressions. There is also a video camera capturing the videos for the facial expression analysis. Our work using video data to analyze facial expressions is beyond the scope of this paper.

## 4. Feature Extraction for Facial Expression Analysis

This section introduces the feature extraction methods we use for facial expression analysis. Our methods automatically detect the fiducial landmarks at 2D images. Since the correspondences between 3D surface data and 2D tex-

(a) (b)

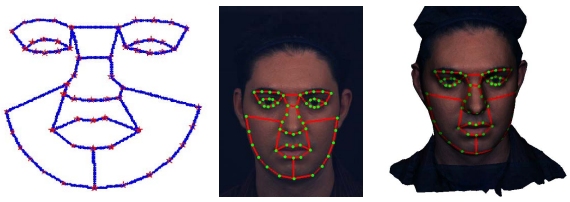
Figure 3. The 3D surface and texture of a participant. (a) surface data. (b) texture mapped on surface data

ture have been established during camera calibration and reconstruction, the landmarks on 2D images also allow defining facial regions at both 2D and 3D data. Based on defined facial regions, important features are extracted from both 3D surface data and 2D textures. The features include the 2D geometric features defined on 2D shapes, the curvature features calculated from the 3D surface data, and the moment invariants which combine both 3D geometry and image texture. The extracted features characterize the facial changes caused by expressions, and are then used to analyze facial expressions of patients with schizophrenia, which will be discussed in Section 5.

### 4.1. Automatic Fiducial Landmarks Detection

Manually labeling fiducial landmarks is usually very time-consuming, and subjective to the person who does the labeling. To automate the process, we first detect faces in images, and then detect fiducial landmarks inside faces, using off-line trained face and fiducial landmark detectors. An AdaBoost based face detector is applied to detect frontal and near-frontal faces [33]. Inside detected faces, our method identifies some important fiducial landmarks using Active Appearance Model (AAM), which is a statistical method to model face appearance as well as face shape [6]. In our implementation that is modified from [27], the face shape is defined by 58 landmarks, as shown in Figure 4.(a). Figure 4.(b) shows the fiducial landmarks detected by AAM. These landmarks are used to define a set of facial regions based on which the features will be extracted for facial expression analysis.

3D facial regions can be automatically created from the 2D shape, since the correspondence between 2D and 3D coordinates has been established during camera calibration and surface reconstruction. Figure 4.(c) shows the landmarks and texture mapped on the surface. Assuming that a 3D point $(x_i, y_i, z_i)$ has its correspondence $(c_i, r_i)$ on a 2D image, then for any facial region $\omega_j = \{(c_{j1}, r_{j1}), (c_{j2}, r_{j2}), ..., \}$ defined on a 2D image, we can define its corresponding 3D facial regions as $\Omega = \{(x_{j1}, y_{j1}, z_{j1}), (x_{j2}, y_{j2}, z_{j2}), ...\}$. The benefit of building the 2D and 3D facial region correspondence is that it avoids the need for registration between two surface since the regions defined on facial landmarks actually represent regions of interests (ROI) across different faces.

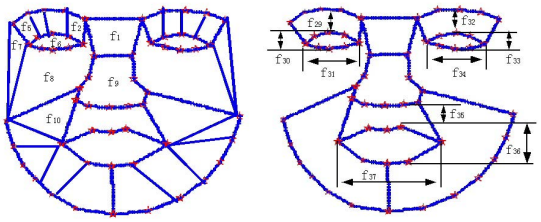


(a) (b) (c)

Figure 4. Fiducial landmarks and 2D/3D correspondence. (a) 58 landmarks defined on face; (b) landmarks detected on an image; (c) landmarks and textures mapped on the surface

### 4.2. 2D Geometric Features

In the acquisition setup, all the faces are at their frontal-view facing the cameras. Therefore the 2D shapes directly characterizes the facial changes. We define two types of 2D geometric features based on the 2D landmarks. One type of 2D geometric features is defined as the area of facial regions to describe the global facial changes. There are totally 28 regions defined on 58 fiducial points, as illustrated in Figure 5.(a). The 28 region areas are denoted as $(f_1^g, f_2^g, ...f_{28}^g)$. There are also some facial actions which are known to be closely related with expression changes. Such facial actions include eye opening, mouth opening and closing, mouth corner movement and eyebrow movement. To describe such actions, we include another type of geometric features, which specifically measure the distance between some fiducial points, as in Figure 5.(b). The distance features are denoted as $(f_{29}^g, ..., f_{37}^g)$. As the combination of two types of feature, the 2D geometric features are denoted as $f^g = (f_1^g, f_2^g, ..., f_{37}^g)$



(a) (b)

Figure 5. Geometric features defined on fiducial landmarks for expression analysis. (a) 28 regions; (b) distance features

### 4.3. 3D Curvature Features

Our method extracts surface curvatures [18] to represent the 3D geometry of faces. The surface curvatures have some desirable properties for facial expression analysis: the curvatures are insensitive to the rigid transformation of surface, therefore the surface does not need accurate registration; the

curvatures will change with the local geometric deformations caused by facial expressions. The surface curvatures have been used in face recognition [12], face detection [5], as well as in facial expression recognition [34].

The 3D surface data is denoted as $(x_i, y_i, z_i), i = 1, ..., n$ where $x_i$, $y_i$ and $z_i$ are the coordinates of $i$-th point at three directions respectively, $n$ is the number of the points on the surface. A continuous method is applied to calculate the curvatures by locally fitting the surface using a biquadratic polynomial approximation. First, each point $(x_i, y_i, z_i)$ at the original surface is re-defined in a local coordinate system by shifting and rotation such that the surface normal at local vertices equals to $(0, 0, 1)$. For simplicity, we also use $(x_i, y_i, z_i)$ to denote the points at the local coordinate system. Then at each local point $(x_i, y_i, z_i)$, the surface can be represented by

$$z_i = a_1^i + a_2^i \tilde{x}_i + a_3^i \tilde{y}_i + a_4^i \tilde{x}_i^2 + a_5^i \tilde{x}_i \tilde{y}_i + a_6^i \tilde{y}_i^2 \quad (1)$$

where $\tilde{x}_i = x - x_i$ and $\tilde{y}_i = y - y_i$. The points $(\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)$ are from the neighboring vertices of the point $(x_i, y_i, z_i)$. The surface fitting using cubic functions provides similar accuracy, but with more computational efforts. Based on the surface fitting, the derivatives are $z_x = a_2$, $z_y = a_3$, $z_{xx} = 2a_4$, $z_{yy} = 2a_6$, and $z_{xy} = a_5$. After surface fitting, the mean curvature $H$ and Gaussian curvature $K$ [18] can be calculated as Eqn (2).

$$\begin{aligned} H(x,y,z) &= \frac{(1+z_y^2)z_{xx} - 2z_x z_y z_{xy} + (1+z_x^2)z_{yy}}{2(1+z_x^2+z_y^2)^{3/2}} \\ K(x,y,z) &= \frac{z_{xx}z_{yy} - z_{xy}^2}{(1+z_x^2+z_y^2)^2} \end{aligned} \quad (2)$$

Based on the sign of Gaussian and mean curvatures, points on the surface can be classified into different shapes, i.e., the HK classification [32]. In our methods, the points are labeled as elliptical concave ($H < 0$ and $K > 0$), saddle ridge ($H > 0$ and $K < 0$), elliptic convex ($H > 0$ and $K > 0$), and saddle valley ($H < 0$ and $K < 0$). In order to characterize surface deformations, we calculate the distribution of each type of HK label at each face region. Assuming that at $j$-th face region, the percentage of $i$-th curvature type is $h_{ij}^c$, the 3D curvature features are then denoted as $f^c$:

$$f^c = (h_{11}^c, ..., h_{ij}^c, ..., h_{4M}^c), (i = 1, ..., 4, j = 1, ..., M) \quad (3)$$

We merge some small regions into bigger regions as in the lower cheek and eyebrow regions, for better estimation of HK label distributions. Finally, we calculate the HK label distributions in 21 regions, i.e., $M = 21$ in Eqn. (3).

### 4.4. Moment Invariants Combining Texture and Geometry

The above 2D and 3D geometric features have not utilized the texture information provided in 2D images. Our method further combines textures with geometric features by calculating moment invariants. Assuming that a function $g(x, y, z)$ is defined on the surface data $(x, y, z)$, the moment $M_{lmn}$ that characterizes the spatial distribution of $g(x, y, z)$ on the surface is defined as:

$$M_{lmn} = \int_x \int_y \int_z x^l y^m z^n g(x, y, z) dx dy dz \quad (4)$$

Here we use the Gabor wavelets as $g(x, y, z)$ in the moment calculation, thus combining the texture information with geometry.

To make the moment features invariant to global rigid transformation, we use 3D moment invariants [21]. The central moments are first computed to get translation invariant features. In central moments, the coordinates are subtracted by the centroid $\bar{x} = \frac{M_{100}}{M_{000}}$, $\bar{y} = \frac{M_{010}}{M_{000}}$, and $\bar{z} = \frac{M_{001}}{M_{000}}$, i.e. $M_{lmn} = \int_x \int_y \int_z (x - \bar{x})^l (y - \bar{y})^m (z - \bar{z})^n g(x, y, z) dx dy dz$. Then three 3D moment invariants, which have been known to be rotation invariant [21], are calculated as Eqn. ( 5).

$$\begin{aligned} J_1 &= M_{200} + M_{020} + M_{002} \\ J_2 &= M_{200}M_{020} + M_{200}M_{002} + M_{020}M_{002} - M_{101}^2 \\ &- M_{110}^2 - M_{011}^2 \\ J_3 &= M_{200}M_{020}M_{002} - M_{002}M_{110}^2 + 2M_{110}M_{101}M_{011} \\ &- M_{020}M_{101}^2 - M_{200}M_{011}^2 \end{aligned} \quad (5)$$

Considering the computational complexity, the Gabor wavelets are only calculated around landmarks, and the Gabor moment invariants are extracted for six facial regions, such as eyes, eyebrows, mouth and nose. At each point, the Gabor wavelets are computed at 3 scales and 6 directions. Therefore, there are total 18 Gabor wavelet coefficients for each point, resulting 18*3 moment invariants for each region. The moment invariants of the $i$-th Gabor coefficient at the $j$-th region are denoted as $J_1^{(i,j)}$, $J_2^{(i,j)}$ and $J_3^{(i,j)}$ respectively. Finally we have the moment invariant features defined as $f^m = (J_1^{(1,1)}, ..., J_1^{(1,M)}, J_2^{(1,1)}, ..., , J_3^{(18,M)})$.

### 4.5. Feature Normalization

To remove the individual differences in the facial expressions, extracted features are normalized in several ways. First, each face shape is normalized to the same scale. We use the width of face (the distance between two ears) for scale normalization, since it won't change with facial expressions. Second, the geometric features are normalized by each subject's neutral faces. For example, each 2D geometric feature is divided by its corresponding value at the neutral expression of the same person. Thus the geometric features only reflect the ratio changes of 2D face geometry, and individual topological differences are canceled. Third, the intensity of face images is normalized by subtracting its

mean and then being divided by its standard deviation, in order to minimize the affects of lighting and skin colors. Finally, all the feature values are normalized to z-scores for further analysis.

## 5. Quantifying Facial Expression Abnormality in Schizophrenia

In this section, we first validate the extracted features via designed facial expression recognition experiments on actors. The actors generally represent expressive people. The validation results demonstrate that classifiers created using our combined features are able to better distinguish different facial expressions than only using 2D or texture features. Then we proceed to quantify facial expression abnormality in schizophrenia. As schizophrenia patients have demonstrated low intensity, flat or inappropriate expressions, traditional facial expression recognition cannot achieves good quantification results for all the expressions. Since we only have a small sample set due to the difficulties of acquiring patient/control data, we instead characterize the group difference of facial expressions in an ISOMAP manifold. For the purpose of quantification, the distances to the control manifold are used for the t-test to provide significance levels for the group differences.

### 5.1. Validation on Actor Database

An actor database, which has been collected during our previous study [13], is used to validate our feature extraction method via facial expression recognition experiments. The actor database contains a total of 32 professional actors. The actors were guided by professional theater directors based on the Russian acting principles through each of the four emotions: anger, fear, happiness, and sadness, at three levels of intensity. Among all the actors, there are 9 actors whose images and surface data have both been acquired. In the following experiments, we use the data of the 9 subjects for validation.

We first cascade all the features together to form a combined feature vector $f = (f^g, f^c, f^m)$. PCA is applied on $f$ to reduce its dimensionality, followed by the LDA to distinguish different facial expressions. Since there are four expressions in the actor database (with no disgust being acquired for the actors), the dimensionality of the LDA subspace equals to 3. We then build a probabilistic K-NN classifier [15] in the LDA subspace. Our experiments show that the probabilistic K-NN classifier actually outperforms the commonly used SVM. The reasons are as follows: the probabilistic K-NN naturally handles multi-class problems while the SVM is essentially a binary classifier; in our case, the SVM needs to optimize the parameters to fit the limited number of training samples, which potentially causes the over-fitting.

We perform a leave-one-sample-out cross validation on the actor database. During the validation, each sample will be left from training once and only once. The validation accuracy is then averaged over all the samples. The validation results are displayed in Table 1. The overall facial expression accuracy is $83.0\%$. We also perform a leave-one-sample-out cross validation, in which each time all the data from one person is left from training. Since we only have totally 9 subjects, the leave-one-subject-out validation is more difficult than leave-one-sample-out validation. As the result, we obtain an overall recognition accuracy of $71.3\%$ for the leave-one-subject-out validation.

Table 1. Confusion matrix of the cross validation on actors

| Expression | Happiness | Anger | Fear | Sadness |
|---|---|---|---|---|
| Happiness | 84.2% | 5.2% | 5.3% | 5.3% |
| Anger | 3.8% | 80.8% | 7.7% | 7.7% |
| Fear | 3.7% | 14.8% | 81.5% | 0% |
| Sadness | 7.4% | 7.4% | 0% | 85.2% |

Also, different features are compared with each other using leave-one-sample-out cross validation. We perform facial expression recognition using only 2D geometric features, using the combined 2D and 3D geometric features, using the Gabor features, and using our combined features. The Gabor features are calculated at 45 landmarks excluding those at the face outline. Table 2 shows the comparison results, which demonstrate that by combining 2D and 3D features, the facial expression analysis accuracy can be largely improved, with our combined features producing the best results.

Table 2. Comparison of different features

| Features | 2D geometric features | 2D geometric + 3D curvatures | Gabor features | Our combined features |
|---|---|---|---|---|
| Average accuracy | 72.4% | 74.5% | 74.0% | 83.0% |

Our validation is further compared with human rating results. In the previous study, 41 students were recruited as human raters from undergraduate and graduate courses in psychology at the Drexel University. The raters were displayed each face, and were asked to identify the emotional content of the face. As the result, the human raters were able to correctly identify 77% correct for anger, 67% correct for fear, 98% for happiness, and 67% for sadness. The overall accuracy of human raters is 77.8%, which is comparable with our cross validation accuracy, i.e., $71.3\%$ for leave-one-subject-out validation and $83.0\%$ for leave-one-sample-out validation.

### 5.2. Quantifying Group Differences between Healthy Controls and Patients in Manifold

The actors used for the facial expression experiments in Section 5.1 are experts in posing expressions. They are in general more expressive than the normal people, and are also more expressive than the healthy controls and patients in our study. This is confirmed by the human rating results showed in Table 3. Students with domain knowledge were asked to rate the evoked expressions of the 12 patients and controls that participated in the study. The faces were rated using the method of FACES [19] by reviewing captured video data. The videos provide rich information, therefore the rating results are more robust than those using single images. However, the human raters can only correctly identify a low percentage (mostly 40% to 70% except for happiness) of intended emotions, for both controls and patients. In such a case, automated facial recognition methods will fail due to the underlying inexpressiveness of individuals. On the other hand, the rating results indeed show differences for controls and schizophrenia patients. We want to quantify potential group differences by applying the extracted features, which have been demonstrated to be able to capture facial expression changes from the previous validations.

Table 3. Accuracy of human rating for evoked expressions

| Expressions | Happiness | Sadness | Anger | Fear | Disgust |
|---|---|---|---|---|---|
| Controls | 89.2% | 39.1% | 71.2% | 63.6% | 55.9% |
| Patients | 65.7% | 57.2% | 53.4% | 50.0% | 53.4% |

As supported by human ratings, the facial expression abnormality in schizophrenia is only demonstrated in a group. Facial expression recognition methods may not be sufficient to distinguish individual patients from controls. Therefore, instead of performing facial expression recognition on patients, we perform a statistical analysis using extract features to quantify the group differences between patients and controls. Our method first embeds facial expression features into a manifold, and then measures the group differences on the manifold. In this method, the ISOMAP manifold learning algorithm is used to obtain a lower dimensional embedding of the original high dimensional features, as it can preserve geodesic distances and provide a non-linear manifold [28]. After obtaining the ISOMAP embedding, the paired t-test is performed to quantify the group differences between healthy controls and patients. In the group difference testing, the distance of each participant to the manifold of healthy controls is calculated as features for the statistical testing. We choose the minimum distance to the control manifold for the testing because the control manifold itself is more cohesive than the patients manifold which could be scattered due to the impaired expression, and controls are used as the baseline for any clinical study.

For the purpose of paired comparison between controls and patients, we applied an out-of-sample ISOMAP extension [7]. This extension specifies landmarks, and only uses the distances to the landmarks for the embedding. In our experiments, each paired control and patient will be left once and only once from the landmarks, and then mean distance $d_i$ from $i$-th participant to the manifold obtained from the remaining healthy controls is computed. Finally, we obtain a set of paired data, i.e., $\{d_1^c, d_2^c, ...\}$ and $\{d_1^p, d_2^p, ...\}$ where $d_i^c$ and $d_i^p$ are distances from $i$-th paired control and patient, for the t-test. Since the pairs have been selected by matching their gender, race and age, the statistical analysis results are not biased by these factors. The t-test is performed for each expression respectively, and the p-values obtained are shown in Table 4.

Table 4. P-values of the t-test for different expressions

| Expressions | Happiness | Sadness | Anger | Fear | Disgust |
|---|---|---|---|---|---|
| Posed | 0.0636 | 0.8246 | 0.1971 | 0.2310 | 0.5668 |
| Evoked | 0.0820 | 0.4636 | 0.4755 | 0.0596 | 0.0247 |

The p-values represent the significance level of the group differences. Table 4 shows the group difference at the significance level around 0.05 in the evoked fear and disgust. The results are consistent with human rating results. Although group difference in other emotions are at a lower significance perhaps due to the small sample size and high variability in the data, we observe trends of difference between the two groups. We expect the results of statistical analysis to improve with more data acquired in the future. Furthermore, the analysis results clearly show that the group differences in evoked expressions are more significant than in posed expressions. Also the emotion of fear shows high group difference between patients and controls. Both of these results are significant in the light of fact that schizophrenia patients show greater impairment in anger and fear, and group differences have clinically been observed better in evoked expressions [13].

## 6. Conclusion

In this paper, we present methods to quantify facial expression abnormality in schizophrenia. We made two major contributions. First, we present a combined feature extraction method, which models both 2D and 3D geometry together with textures. The features are demonstrated to provide better accuracy in actors' recognizing facial expressions. Second, we analyze the facial expression abnormality of schizophrenia patients. The feature distance of participants to the healthy control manifold are used for the paired t-test. Our statistical analysis results are well correlated with human ratings and clinical findings that schizophrenia patients have impaired facial expression. Our method is general and applicable to the further study of other clinical conditions that cause deficits in expressiveness such as

autism and Parkinson's disease. The future work will continue collecting more paired participants (patients and controls) for the statistical analysis. We also plan to combine the 3D data with the video data for the facial expression analysis, as well as creating point-wise expression change maps for further improvement.

## Acknowledge

We would like to acknowledge support from NIH grants 1R01MH73174-01 (for method development) and R01-MH060722 (for data acquisition).

## References

[1] M.S. Bartlet, J. C. Hager, P. Ekman, and T. J. Sejnowski, *Measuring facial expressions by computer image analysis*, Psychophysiology **36** (1999), 253–263.

[2] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J. Movellan, *Recognizing facial expression: Machine learning and application to spontaneous behavior*, CVPR, 2005, pp. 568–573.

[3] Y. Chang, C. Hu, and M. Turk, *Probabilistic expression analysis on manifolds*, CVPR, vol. 2, 2004, pp. 520–527.

[4] I. Cohen, N. Seve, G. G. Cozman, M. C. Cirelo, and T. S. Huang, *Learning Bayesian network classifier for facial expression recognition using both labeled and unlabeled data*, CVPR, vol. 1, 2003, p. 595.

[5] A. Colombo, C. Cusano, and R. Schettini, *3D face detection using curvature analysis*, Pattern Recognition **39** (2006), no. 3, 444–455.

[6] T.F. Cootes, G.J.Edwards, and C.J. Taylor, *Active appearance models*, IEEE. Trans. on PAMI **23** (2001), no. 6, 681–685.

[7] V. de Silva and J. B. Tenenbaum, *Global versus local methods in nonlinear dimensionality reduction*, Advances in Neural Information Processing Systems, vol. 15, Cambridge, MIT Press, 2002, pp. 705–712.

[8] P. Ekman and W. V. Friesen, *Facial action coding system: A technique for the measurement of facial movement*, Consulting Psychologists Press, Palo Alto, Calif., 1978.

[9] I.A. Essa and A.P. Pentland, *Facial expression recognition using a dynamic model and motion energy*, ICCV, 1995, pp. 360–367.

[10] B. Fasel and J. Luettin, *Recognition of asymmetric facial action unit activities and intensities*, ICPR, vol. 1, 2000, pp. 1100–1103.

[11] ———, *Automatic facial expression analysis: a survey*, Pattern Recognition **36** (2003), no. 1, 259–275.

[12] G. G. Gordon, *Face recognition based on depth and curvature features*, CVPR, 1992, pp. 808–810.

[13] R. C. Gur, R. Sara, and M. Hagendoorn et. al, *A method for obtaining 3-dimensional facial expressions and its standardization for the use in neurocognitive studies*, Journal of Neuroscience Methods **115** (2002), no. 2, 137–143.

[14] R. E. Gur, C. G. Kohler, J. D. Ragland, S. J. Siegel, K. Lesko, W. B. Bilker, and R. C. Gur, *Flat affect in schizophrenia: relation to emotion processing and neurocoginitive measures*, Schizophrenia Bulletin **32** (2006), no. 2, 279–287.

[15] C. C. Holmes and N. M. Adams, *A probabilistic nearest neighbor method for statistical pattern recognition*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **64** (2002), 295–306.

[16] T. Kanade, J.F. Cohn, and Y. Tian, *Comprehensive database for facial expression analysis*, AFRG, 2000, pp. 46–53.

[17] I. Kotsia and I. Pitas, *Real time facial expression recognition from image sequences using support vector machines*, ICIP, 2005, pp. 966–969.

[18] Erwin Kreyszig, *Differential geometry*, Dover Pubns, 1991.

[19] A. M. Kring, M. Alpert, J.M. Neale, and P.D. Harvey, *A multichannel, multimethod assessment of affective flattening in schizophrenia*, Psychiatry Research **54** (1994), 211–222.

[20] J. J. Lien, T. Kanade, J.F. Cohn, and C-C Li, *Detection, tracking, and classification of action units in facial expression*, Robtics and Autonomous Systems **31** (2000), 131–146.

[21] C.-H. Lo and H.-S. Don, *3-D moment forms: their construction and application to object identification and positioning*, IEEE. Trans. on PAMI **11** (1989), no. 10, 1053 – 1064.

[22] M. J. Lyons, J. Budynek, and S. Akamatsu, *Automatic classification of single facial images*, IEEE. Trans. on PAMI **21** (1999), no. 12, 1357–1362.

[23] M. Pantic and L.J.M. Rothkrantz, *Automatic analysis of facial expressions: the state of the art*, IEEE. Trans. on PAMI **22** (2000), no. 12, 1424–1445.

[24] ———, *An expert system for recognition of facial actions and their intensity*, Int'l Conf. on Innovative Applications of Artificial Intelligence, 2000, pp. 1026–1033.

[25] J. E. Salem and A. M. Kring, *More evidence for generalized poor performance in facial emotion perception in schizophrenia*, Journal of Abnormal Psychology **105** (1996), no. 3, 480–483.

[26] R. Sara and R. Bajcsy, *Fish-scales: representing fuzzy manifolds*, ICCV, 1998, pp. 811–817.

[27] M. B. Stegmann, B. K. Ersbll, and R. Larsen, *FAME - a flexible appearance modelling environment*, IEEE Trans. on Medical Imaging **22** (2003), no. 10, 1319–1331.

[28] J. B. Tenenbaum, V. de Silva, and J. C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science **290** (2000), no. 5500.

[29] Y-L Tian, T. Kanade, and J. F. Cohn, *Handbook of face recognition*, ch. Facial Expression Analysis, pp. 247–275, Springer, 2005.

[30] Y-L Tian, T. Kanade, and J.F. Cohn, *Recognizing action units for facial expression analysis*, IEEE. Trans. on PAMI **23** (2001), no. 2, 97–115.

[31] Y. Tong, W. Liao, and Q. Ji, *Inferring facial action units with causal relations*, CVPR, vol. 2, 2006, pp. 1623–1630.

[32] E. Trucc and R. B. Fisher, *Experiments in curvature-based segmentation of range data*, IEEE. Trans. on PAMI **17** (1995), no. 2, 177–182.

[33] P. Viola and M. Jones, *Robust real-time object detection*, International Journal of Computer Vision **57** (2004), no. 2, 137–154.

[34] J. Wang, L. Yin, X. Wei, and Y. Sun, *3D facial expression recognition based on primitive surface feature distribution*, CVPR, vol. 2, 2006, pp. 1399 – 1406),.

[35] Y. Yacoob and L.S. Davis, *Recognizing human facial expressions from long image sequences using optical flow*, IEEE. Trans. on PAMI **18** (1996), no. 6, 636–642.

[36] M. Yeasin, B. Bullot, and R. Sharma, *From facial expression to level of interest: a spatio-temporal approach*, CVPR, vol. 2, 2004, pp. 922–927.

[37] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, *A 3D facial expression database for facial behavior research*, AFRG, 2006, pp. 211–216),.

[38] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, *Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron*, AFRG, 1998, pp. 454–459.