Human Emotion Recognition Using Real 3D Visual Features from Gabor Library

Tie Yun #1, Ling Guan#2

# Ryerson Multimedia Research Lab
Ryerson University, Toronto

1 ytie@ee.ryerson.ca
2 lguan@ee.ryerson.ca

Abstract—Emotional state recognition is an important component for efficient human-computer interaction. Most existing works address this problem using 2D features, but they are sensitive to head pose, clutter, and variations in lighting conditions. The general 3D based methods only consider geometric information for feature extraction. In this paper, we present a real 3D visual features based method for human emotion recognition. 3D geometric information plus colour/density information of the facial expressions are extracted by 3D Gabor library to construct visual feature vectors. The filter’s scale, orientation, and shape of the library are specified according to the appearance patterns of the 3D facial expressions. An improved kernel canonical correlation analysis (IKCCA) algorithm is proposed for final decision. From training samples, the semantic ratings that describe the different facial expressions are computed by IKCCA to generate a seven dimensional semantic expression vector. It is applied for learning the correlation with different testing samples. According to this correlation, we estimate the associated expression vector and perform expression classification. From experiment results, our proposed method demonstrates impressive performance.

I. INTRODUCTION

Emotion plays a critical role in human-to-human interaction and communication, allowing people to express oneself beyond the verbal domain and understand to each other with the assistance of the contextual information from various modalities. The ability to recognize human affective or emotional state is desirable to empower an intelligent computer to interpret, understand, and respond to human emotions, moods, and, possibly, intentions, which are similar to the way that humans rely on their senses to assess each other’s affective state [1]. Many potential applications such as intelligent automobile systems, game and entertainment industries, interactive video, indexing and retrieval within large scale image and/or video databases can benefit from this ability.

The vision based works mainly focus on facial expression analysis because of the importance of face in emotion expression and perception [2]. In spite of considerable previous works documented in this area, some challenging are still remained. The majority approaches for solving human emotion recognition problem attempt in identification on 2D spatiotemporal data, either 2D static images or 2D video sequences. They are sensitive to head poses, clutter, and variations in lighting conditions. The general 3D approaches only use geometric information without any colour/density information of the face. However, these features can strongly affect the performance of an emotion recognition system [3].

In this paper, we present a new emotion recognition method using 3D geometric features with facial density information. The main contribution of this work consists of applying the 3D Gabor library for feature extraction and an IKCCA algorithm for final classification. To the best of our knowledge, there is no same work has been done with 3D Gabor library for emotion or face recognition. The proposed work is the first try using such visual features for face analysis. The block diagram of this method is shown in Fig. 1. We extract primitive 3D facial expression feature vectors by the 3D Gabor library. Then the IKCCA is applied for seven dimensional semantic expression vector ratings and for classifying the prototypic facial expressions with the trained facial feature distribution. To validate our proposed approach, we have conducted experiments for person-independent facial expression recognition on a public 3D facial expression database, i.e. the BU_3DFE database [4]. The experimental results demonstrate the performance of 85.39% as the overall recognition rate.

The rest of the paper is organized as follows. Section 2 summarizes the related works. In Section 3, the 3D Gabor for feature extraction is described. We present the IKCCA algorithm in Section 4. The experimental results and conclusions are given in Section 5 and 6, respectively.
II. RELATED WORKS

Many of the existing human emotion recognition systems attempt to recognize some prototypic emotions. The most important and widely accepted set of measures is composed of happiness, sadness, anger, fear, surprise, and disgust, which was proposed by Ekman and Friesen [5]. These six emotions are not culturally determined, but universal to human culture and thus biological in nature. The psychological studies indicated that showing facial expression is a natural and primary way for communicating the quality and nature of emotions. Each of the six emotions corresponds to a unique facial expression.

Different methods have been explored to perform facial expression analysis which can be roughly categorized into two groups, i.e. the geometric feature-based methods and the appearance-based methods. Silva and Hui [6] determine the eye and lip position using low-pass filtering and edge detection method. They achieved an average emotion recognition rate of 60% using a neural network (NN). Anderson and McOwan [7] presented an automated multistage system for real-time recognition of facial expression. The system used facial motion to characterize monochrome frontal views of facial expressions and was able to operate effectively in cluttered and dynamic scenes, recognizing the six emotions universally associated with unique facial expressions. Lyons et al. [8] use a set of multi-scale, multi-orientation Gabor filters to transform the images. They tested their system with a database of 193 images posed by 9 Japanese females, and achieved an expression classification accuracy of 75%. Wang and Guan [9] constructed a bimodal system for emotion recognition. They use a face detection scheme based on the HSV color model to detect the face from the background and Gabor wavelet features to represent the facial expressions. They achieved the best overall recognition rate of 82.14% using a multi-classifier scheme.

The performance of 2D based algorithms remains unsatisfactory, and is often unreliable under adverse conditions. It is difficult to handle pose variations, lighting illumination and subtle facial behavior. Therefore, the aforementioned methods are limited to constrained environment. To achieve more robust performance, a growing body of research has been focused on addressing the problem using 3D information. The analysis on 3D facial expressions facilitates the examination of the fine structural changes inherent in the spontaneous expressions. The 3D based algorithm allows the transfer of feature-like models from the given single view into arbitrary views, thus making the solution far more pose invariant than the current 2D solutions. Furthermore, 3D data are by definition illumination invariant, thus eliminating errors associated with changes in illumination becomes more practical with knowledge of the physical surfaces being considered.

Song et al. [10] presented a generic facial expression analogy technique to transfer facial expressions between arbitrary 3D face models, as well as between 2D face images. Geometry encoding for triangle meshes, vertex-tent-coordinates were proposed to formulate expression transfer in 2D and 3D cases as a solution to a simple system of linear equations. Chin et al. [11] presented an emotional intensity-based facial expression modeling process by generating a 3D customized face and facial expressions. The generated customized face integrates expression data using different expression intensities. They identified six universal expressions by determining the anatomical, parametric values of linear and sphincter muscles. Facial expressions were also simulated by intensity mapping. The best recognition rates for 3D data analysis are achieved by Tang and Huang [12]. Using a regularized multi-class AdaBoost classification algorithm, they achieved a 95.1% average recognition rate for six universal facial expressions with a highest average recognition rate of 99.2% for the recognition of surprise. However, this method is a feature point based method and the positions have to be located manually. Our method is appearance-based method, deal with 3D data directly, no human interaction involved, thus is a more general approach for 3D face processing.

III. GABOR FEATURE EXTRACTION

Gabor transform based facial expression recognition systems show that their representation has a high degree of correlation with the human semantic ratings. 3D Gabor transform for feature extraction can accurately describe the particular set of facial expression to discriminate between expressions.

A. 3D Gabor Filter

A 3D Gabor transform is basically a product of a complex sinusoid wave modulated by a 3D Gaussian window. We have the 3D Gabor transform as the following:

\[
G_{y,x,\beta}(x',y',z') = A \times H(x',y',z') \times S(x,y,z)
\]

where \( A = \frac{1}{(2\pi)^{3/2}} \sigma_x \sigma_y \sigma_z \) is a normalization scale, and

\[
H(x',y',z') = \exp\left\{-\frac{1}{2} \left( \left( \frac{x'}{\sigma_x} \right)^2 + \left( \frac{y'}{\sigma_y} \right)^2 + \left( \frac{z'}{\sigma_z} \right)^2 \right) \right\}
\]

\[
S(x,y,z) = \exp\left[-j2\pi(Mx+Ny+Lz)\right]
\]

\((x',y',z')\) is non-rotated spatial coordinates, \( \sigma_x, \sigma_y, \sigma_z \), define the width of the Gaussian envelop along x, y and z axes respectively. \((x',y',z')^T = R \times (x,y,z)^T\) is the rotated spatial coordinates of the 3-D Gaussian envelop, \( R \) is a rotation matrix for transforming the Gaussian envelop to coincide with orientation of the sinusoid. We also have \( N = \gamma \sin \beta \cos \alpha, M = \gamma \sin \beta \cos \alpha, \) and \( L = \gamma \cos \beta \) that are calculated from the 3D frequencies, denoted by \( \alpha, \beta \) and \( \gamma \) of the complex sinusoid.

Fig. 2 shows the projections of a 3D Gabor filter with \( \sigma_x = \sigma_y = \sigma_z = \sigma, \gamma = 0.25, \alpha = \pi/2, \beta = \pi/2, \sigma = 1/\gamma \), where the size of the filter is 60x60x60.
B. Gabor Library Design

Since the prior information about the 3D facial expression is unknown, we design the 3D Gabor library using a set of Gabor filters with different frequencies and orientations to obtain sufficient information as the following:

\[
G_{k,v,w}(x,y,z) = \frac{1}{(2\pi)^{2}} \sigma_{\alpha} \sigma_{\beta} e^{-\frac{1}{2} \left( \frac{x^{2}}{\sigma_{\alpha}^{2}} + \frac{y^{2}}{\sigma_{\beta}^{2}} \right)} e^{-j2\pi(Mv+Nz+Lz)}
\]

where \(\gamma_{\text{max}}\) is the upper centre frequencies of the signal to be analysed. We denote the wavelets as

\[
\{ G_{k,v,w}; k=0,...,K-1, v=0,...,V-1, w=0,...,W-1 \}
\]

The feature of a 3D facial expression can be extracted with the frequency and orientation information from the voxel of the Gabor library. Since the magnitude of the convolution result can express the response of a wavelet to the facial expression, the useful information of intensity changes at the face can be obtained by applying the Gabor library on a 3D facial expression. In this paper, the library parameters are set to \(4 \times 4 \times 4\). In Fig. 3, we show the designed 3D Gabor filters library in both space domain and frequency domain.

C. Feature Representation

In this subsection, we describe our method for visual feature extraction using the 3D Gabor library. We normalize all the input data to be the same size and denote it as \(f(x,y,z)\). The Gabor transform \(F(x,y,z)\) of this volume \(I(x,y,z)\) can be calculated as:

\[
F_{k,v,w}(x,y,z) = \int f(x,y,z) G_{k,v,w}(x-a_{x},y-b_{y},z-c_{z}) dx dy dz
\]

where * indicates the complex conjugate. In this case, we obtain a very huge coefficient matrix for each face. We use a total of \(4 \times 4 \times 4 = 64\) Gabor filters, and thus the size of the matrix is \(60 \times 60 \times 60 \times 64\). With a feature space of such high dimensionality, the computation cost is very high, and thus it is not suitable for fast processing. We, hence, take the mean and standard deviation of the magnitude of the transform coefficients of each sub-band filter as the features:

\[
\mu_{k,v,w} = \frac{1}{\sigma_{k,v,w}^{2}} \int \int \int |F_{k,v,w}(x,y,z)|^{2} dx dy dz
\]

Then we construct a feature vector for each emotional face using \(\mu_{k,v,w}\) and \(\sigma_{k,v,w}\) as feature components. In this paper, we use four scales \(K=4\) and four orientations \(V=W=4\) in \(\alpha\) and \(\beta\) directions, resulting in a feature vector of 128 dimensions for classification, which can be written as:

\[
\{ \mu_{000}, \sigma_{000}, \mu_{001}, \sigma_{001}, ..., \mu_{333}, \sigma_{333} \}
\]

IV. IKCCA CLASSIFIER

The IKCCA classifier is based on learning the maximum correlation between a test feature vector and a semantic expression vector. We propose to adopt the IKCCA method for emotion recognition. This method depends on normalized cross covariance and has been theoretically proven that it solves the singularity problem of the general KCCA algorithm, which cannot be accomplished by using the regularization method or the eigenvalue decomposition method.

Let \(X \in R^{n}\) be the test feature vector and \(Y \in R^{n}\) be the semantic expression vector. \(A = a^{T}X, B = b^{T}Y\) are their projections with zero mean. To determine the corresponding emotional state for a given vector \(X\), we need to find a pair of directions \(a(X)\) and \(b(Y)\) so the canonical correlation is maximized:

\[
\rho(X,Y;a,b) = \frac{E[AB]}{\sqrt{E[A^{2}]E[B^{2}]}} = \frac{E[a_{x}^{T}XYb_{y}]}{\sqrt{E[a_{x}XX^{T}a_{x}]E[b_{y}YY^{T}b_{y}]}}
\]

Let \(\Phi(X)\) and \(\Psi(Y)\) denote the diagonals of \(X\) and \(Y\) in Hilbert space through nonlinear mapping respectively. The correlation function \(\rho\) can be reformulated as:

\[
\rho(\Phi(X),\Psi(Y);a_{\Phi(X)},b_{\Psi(Y)}) = \frac{\text{Cov}[\Phi(X),\Psi(Y)]}{\sqrt{\text{Var}[\Phi(X)]\text{Var}[\Psi(Y)]}}
\]

where

\[
\text{Cov}[\Phi(X),\Psi(Y)] = \frac{1}{n} \sum_{i=1}^{n} (\Phi(X)_{i} - \mu_{\Phi(X)}) (\Psi(Y)_{i} - \mu_{\Psi(Y)})
\]

and

\[
\text{Var}[\Phi(X)] = \frac{1}{n} \sum_{i=1}^{n} (\Phi(X)_{i} - \mu_{\Phi(X)})^{2}
\]

507
where $\text{Cov}[\Phi(X), \Psi(Y)]$ is the empirical covariance. For classifying human’s emotional state, we need to calculate the empirical solution to (10). We use a positive constant factor $\varepsilon_\pi$ in the calculation of the overall correlation $\rho$ so that the objective function becomes:

$$\rho = \frac{1}{n} \sum_{i=1}^{n} \left( \Phi_i(X) - \frac{1}{n} \sum_{i=1}^{n} \Phi_i(X) \right) \left( \Psi_i(Y) - \frac{1}{n} \sum_{i=1}^{n} \Psi_i(Y) \right)$$

$$\sqrt{\sum_{i=1}^{n} \Phi_i(X)^T \Phi_i(X) + \sum_{i=1}^{n} \Psi_i(Y)^T \Psi_i(Y)}$$

Using IKCCA for the classification, first we set $\Phi(X_{test})$ to be an input test feature vector and $\Psi(Y_{semantic})$ be the corresponding semantic expression vector. Let $A_{test}$ be the projection of $\Phi(X_{test})$ onto the directions $a_i \Phi(X)$ and $B_{semantic}$ be the projection of $\Psi(Y_{semantic})$ onto the directions $b_i \Psi(Y)$ so we have:

$$A_{test} = [A_{test}^1, \ldots, A_{test}^n] = [a_1 \Phi(X), \ldots, a_n \Phi(X)]^T \Phi(X_{test}) = P_x^T \Phi(X_{test})$$

$$B_{semantic} = [B_{semantic}^1, \ldots, B_{semantic}^n] = [b_1 \Psi(Y), \ldots, b_n \Psi(Y)]^T \Psi(Y_{semantic}) = P_y^T \Psi(Y_{semantic})$$

Using (11), (12) and (13), we can estimate the corresponding semantic expression vector $Y_{semantic}^i$ for a given input feature vector $X_{test}^i$. The index of the best matched emotion class of the input sample is then given by:

$$C^* = \arg \max_{i=1}^{n} \left( P_x^T \Phi(X_{test}^i), P_y^T \Psi(Y_{semantic}^i) \right)$$

$$= \arg \max_{i=1}^{n} \rho \left( X_{test}^i, Y_{semantic}^i; a_i \Phi(X), b_i \Psi(Y) \right)$$

V. EXPERIMENT AND RESULTS

To evaluate the performance of the proposed emotion recognition method, we implement the experiments in MatLab on a Pentium IV 2.8 GHz PC with 3.25 GB memory, under the Windows XP operating system.

A. Facial Expression Database

The BU_3DFE database is used in our experiment to validate our proposed approach. The BU_3DFE database is the largest publicly available dataset for 3D facial expression recognition research and contains images exhibiting substantial expression variations, which can cause problems for many recognition algorithms.

In the BU_3DFE 3D database, there are 100 subjects who participated in face scans, including undergraduates, graduates and faculty from State University of New York at Binghamton. The resulting database consists of about 60% female and 40% male subjects with a variety of ethnic/racial ancestries. Each subject in the database performed seven expressions (including neutral), captured by a 3D face scanner. From the database, we choose 560 samples for classifier training and 280 samples for testing, and each delivers one of the seven facial expressions. There is no overlap between the training and testing subjects.

B. Feature Selection

The performance of the IKCCA based emotion recognition system depends on how to find out the best correlations between the feature vectors for the classification task. For a pattern recognition system, the length of the feature vector and the discriminating ability of the features in terms of separating patterns belonging to different classes in the feature space will critically affect the overall performance of the system. The importance of selecting relevant subset from the original feature set is closely related to the “curse of dimensionality” problem in function approximation, in which sample data points become increasingly sparse as the dimensionality of the function domain increases. The finite set of samples may not be adequate for characterizing the original mapping and the computational requirement is higher for implementing a high dimensional mapping.

In this paper, we use Principal Component Analysis (PCA) to reduce the dimensionality of the input feature vector to alleviate the aforementioned problems by reducing the number of transformed features, whilst retaining most of the intrinsic information content of the original data. PCA technique arranges the feature vector in descending order of variance and is truncated at desired length such that the remaining feature vector is sufficient for accurate recognition of facial expressions.

In PCA parameter determination, we first produce the data set with zero mean, and then compute the covariance matrix from the covariance between two dimensions. New data vectors are formed by projecting the original data onto the principal component vectors with the following criterion. The components of the new feature vector are derived from the first $M$ eigenvectors that satisfy $\sum_{i=1}^{M} \lambda_i$ and $\lambda_i \geq 0.9$ ,

C. IKCCA Classifier

To clarify the relationship between feature vectors and the corresponding emotion categories, in our experiment, all kinds of expression feature vectors of seven facial expressions collected from different people are extracted and compared using individual IKCCA classifiers. We show the recognition rates by individual IKCCA classifiers with and without PCA feature selection in Table I. From the table we can see that by applying PCA, the system performance is improved.
The IKCCA is carried out on the training samples of the 3D Gabor feature vectors, and then used to classify the data. This technique attempts to maximize the intra-class correlation following (11) and (14), whilst minimizing the overall correlations between the projected means of different samples we intend to separate. In Table II, we show the confusion matrix of the IKCCA classifier on the BU_3DFE database. From Table II, we can see that features representing different expressions exhibit great diversity since the correlations between different emotions are relatively low. On the other hand, the same expressions collected from different subjects are very similar due to the fact that they are highly correlated within the same class. The overall recognition rate using individual classifier with PCA is about 73.64%.

The confusion matrix also illustrates the most common misclassifications. In general, emotions of sadness, fear and disgust have low recognition rates, as they do not occur naturally alone but associated with other emotions. For example, in the experiment, sadness is often confused with fear. In reality, these two expressions are often blended and facial samples labeled as sadness may include some visual patterns representing the expression of fear. Moreover, the misclassifications can be attributed to the confusion inherently associated with the limited categories of expressions being studied.

We compare the 3D Gabor based IKCCA classifier with and without PCA selection, and a 2D Gabor based method by [9]. Fig. 4 summarizes the results of the best classification accuracy as an average percentage achieved by these methods. From the figure, it is very clearly that when classifying the seven prototypic facial expressions, using the 3D Gabor feature with PCA selection representation obtains a recognition rate of 73.64%, that significantly exceeds the recognition rate of 49.29% by 2D Gabor based method and 67.3% without PCA selection.

D. Semantic Ratings Classification

To further improve the recognition rate, we introduce a semantic rating classification method based on the individual IKCCA classifier by considering the overall correlations for the different classes. We use the calculated semantic ratings for quantitatively evaluating the seven emotional expressions. The semantic ratings are computed from the training samples and combined into a seven dimensional semantic expression vector for expression analysis. For a new 3D facial expression query, we first generate the feature vector using the 3D Gabor filters. The corresponding semantic ratings of each facial expression are estimated using (9), and then the emotion classification is performed according to (14). Fig. 5 illustrates some samples of the semantic ratings for the facial expression model.

From the results we find that expressions with a negative value correspond to a negative reaction in terms of arousal and stance, while positive values correspond to a positive reaction. Note, the expressions that are detected with high accuracy and low confusion are in the classed of happiness, anger and surprise. The reason is that, in general, these emotions have strong negative or positive values with more distinguishable corresponding facial expressions.
We then use both the “one-face-against-all” and “leave-one-subject-out” cross validation strategies to improve the experiment. In the “one-face-against-all” strategy, one facial expression model is used as the testing data and the others as the training data. The process is repeated for all the possible trials. In the “leave-one-subject-out” strategy, the facial expressions belonging to one subject are used as the testing data and the remainders as the training data. This is repeated for all the possible trials until all the subjects are used as the testing data. The overall recognition rate can achieve to 85.39% using the combination of these two strategies.

![Fig. 6. Final Emotion Recognition Rate](image)

The experiment results are averaged to produce the final recognition rate and shown in Fig. 6. From Fig. 6, we observe a total 12% overall performance improvement by the combination of these two strategies compared with the individual classifier.

### TABLE III

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dim</td>
<td>2D</td>
<td>2D</td>
<td>3D</td>
<td>3D</td>
</tr>
<tr>
<td>Fea</td>
<td>geometric</td>
<td>Gabor</td>
<td>geometric</td>
<td>Gabor</td>
</tr>
<tr>
<td>INT.</td>
<td>Auto</td>
<td>Auto</td>
<td>manually</td>
<td>Auto</td>
</tr>
<tr>
<td>Clas</td>
<td>NN</td>
<td>FLDA</td>
<td>AdaBoost</td>
<td>IKCCA</td>
</tr>
<tr>
<td>R. R</td>
<td>60%</td>
<td>49.29%</td>
<td>95%</td>
<td>85.39%</td>
</tr>
</tbody>
</table>

A comparison of the recognition rates achieved by several different methods is depicted in Table III. It shows that the proposed method has achieved the second best detection rate among the recent state-of-the-art methods. The best method has been the one proposed in [12], where a total 95% recognition rate has been achieved using a method based on 3D geometric feature analysis. The main drawback of the method in [12] is that it is only tested in perfect manually aligned feature points’ positions and no experiments in fully automatic conditions have been presented. The propose method has advantages at automatically initializing and real 3D visual feature generating with acceptable recognition rates. In addition, the proposed method has demonstrated its ability to handle pose variations and lighting illumination problems.

### VI. CONCLUSIONS

In this paper, we proposed a robust emotion recognition method using a 3D facial model. The main contribution is that we constructed a 3D Gabor library for facial feature extraction with IKCCA algorithm for final decision. From this work we found that using 3D data for emotion analysis has clear advantage at handling problems such as pose variations, lighting illumination and subtle facial behavior. We also applied the IKCCA algorithm for seven dimensional semantic expression vector ratings to classify the prototypic facial expressions. We observed superior performance by the proposed method when compared with several other methods.

### REFERENCES