Turk J Elec Eng & Comp Sci, Vol.18, No.6, 2010, © TÜBİTAK
doi:10.3906/elk-0908-158

Optimal feature selection for 3D facial expression recognition using coarse-to-fine classification

Hamit SOYEL[1], **Hasan DEMİREL**[2]
[1]*Department of Computer Engineering, Cyprus International University, Nicosia-North Cyprus, Via Mersin 10, TURKEY*
e-mail: hsoyel@ciu.edu.tr
[2]*Department of Electrical and Electronic Engineering, Eastern Mediterranean University, Famagusota-North, Cyprus, via Mersin 10, TURKEY*

Abstract

Automatic facial expression recognition for novel individuals from 3D face data is a challenging task in pattern analysis. This paper describes a feature selection process for pose-invariant 3D facial expression recognition. The process provides a lower dimensional subspace representation, which is optimized to improve the classification accuracy, retrieved from geometrical localization of facial feature points to classify facial expressions. Fisher criterion-based approach is adopted to provide a basis for the optimal selection of features. Two-stage probabilistic neural network architecture is employed as a classifier to recognize the facial expressions. In the first stage, which can be regarded as the coarse classification, the facial expressions are classified into one of the three expression groups formed using seven basic facial expressions. In the fine classification stage, final expression is determined by using within group classification. Facial expressions such as Neutral, Anger, Disgust, Fear, Happiness, Sadness, and Surprise are successfully recognized with an average recognition rate of 93.72%.

1. Introduction

During the past two decades, human expression recognition has attracted a significant interest in the pattern recognition and artificial intelligence, as it plays a vital role in human-computer interaction. Many applications, such as emotion analysis, virtual reality, video-conference, medical nursing, and customer satisfaction studies for shop and restaurant services and so on, require efficient human expression recognition in order to achieve the desired results. Therefore, the impact of human expression recognition on the above-mentioned application areas is constantly growing. People easily distinguish expressions. But, it is a very defiant task because human expression zis dependent on so many factors, including age, race, sex, illumination and so forth. Due to it dynamic structure, it is hard to precisely model the face with global parameters. The pioneering studies of human facial expressions introduced by Ekman [1] gave evidence to the classification of basic facial expressions as per happiness, sadness, anger, fear, surprise, disgust and neutral. Ekman and Friesen [2] developed the Facial

Action Coding System to code facial expressions in which the movements on the face are described by action units. This work inspired many researchers to analyze facial expressions in 2D by means of image and video processing, where by tracking facial features and measuring the amount of facial movements, they attempt to classify different facial expressions.

Many facial expression recognition researchers are focused on visible spectrum images, such as intensity or color images of faces, and have shown reasonable performance under controlled inner and outer environments. Yet, there are still many unsolved problems in applications with variable environments such as those involving pose and illumination changes. With the development of 3D acquisition systems, 3D face capture is becoming faster and cheaper. Facial feature recognition based on 3D information is attracting great interest in order to solve drawbacks of 2D approaches.

As far as the classification of expressions from 3D face data is concerned, Wang et al. in [3] assume isometric properties of skin during deformation by extracting 12 primitive facial surface features from 7 expressive regions, with analytics based on the principal curvature information estimated from the 3D triangle mesh model as given in the BU-3DFE database [4].

Wang and Yin used principal component analysis (PCA) and linear discriminant analysis (LDA) for classification and claim an average person-independent expression recognition rate of 83.6%, which is, according to the authors, better than that realized by the 2D-image-based methods [5]. On the same database, Tang and Huang [6] extracted 96 normalized distances and slopes of line segments connecting 3D facial points as features, invoked a multi-class support vector machine (SVM) classifier, and claimed 87.1% average recognition rate, which is better than their own accuracy of 83.6% using LDA [7] . Recently, Mpiperis et al. [8] proposed bilinear models for joint identity and expression recognition while claiming a recognition accuracy of 90.5% on the BU-3DFE database.

In this paper, we propose to construct subspaces, which are optimized for 3D facial expression classification. One of the major contributions of this work is to analyze facial expressions in 3D space by exploring the facial distance vectors. The distance measures extracted from the 3D facial features provide reliable and valuable information for robust recognition of facial features. Especially, the 3D facial features can be used to correct the pose corresponding 2D facial image and eliminate the interference of illumination.

We also propose a decision-tree based probabilistic neural network (PNN) classification under a coarse-to-fine scheme. This process is composed of two stages of PNN classification. Due to their structural similarities represented by the Mahalanobis distance between the 7 basic expression classes in fisher discriminant space, 3 overlapping clusters referred as the class groups have been formed.

Group 1 contains Surprise; Group 2 contains Anger, Sadness and Neutral; and Group 3 contains Happy, Disgust and Fear. In the coarse classification stage, the cluster group of the query expression is determined with a PNN. Then, in the fine classification stage, for cluster groups 2 and 3, a dedicated PNN is employed to perform the final classification. The PNNs are trained by an iterative selection of individual features that are more salient at each stage. The proposed fisher criterion based feature selection process with PNN generated 88.5% facial expression recognition performance. The addition of coarse-to-fine approach, in the form of tree-PNN, has increased the overall facial expression recognition performance to 93.7%.

Organization of this paper is as follows. Section 2 introduces methodology used for feature selection from a statistical analysis. Construction of a decision tree classifier, based on a coarse-to-fine classification approach and experimental results are dealt in Section 3. Concluding remarks are presented in Section 4.

2. Feature selection methodology

In attempting to classify real-world objects or concepts using computational methods, the selection of an appropriate representation is of considerable importance. For facial expression recognition, the patterns are generally represented as a vector of feature values. The problem of dimensionality reduction encompasses both feature selection and feature extraction. Feature extraction is the process of deriving new features from the original features in order to reduce the cost of feature measurement, increase classifier efficiency, and allow higher classification accuracy. The selection of features can have a considerable impact on the effectiveness of the resulting classification algorithm. It is not often known in advance which features will provide the best discrimination between classes, and it is usually not feasible to measure and represent all possible features of the objects being classified. As a result, feature selection and extraction methods have become important techniques for automated pattern recognition. The main purpose of feature selection is to reduce the number of features used in classification while maintaining acceptable classification accuracy. Less discriminatory features are eliminated, leaving a subset of the original features which retains sufficient information to discriminate well among classes.

In the adopted dimensionality reduction process the 3D distance vectors representing facial expressions (Section 2.1) are transformed into an eigenspace where the basis of the space is determined through the proposed selection process, explained in Section 2.2.

2.1. Facial feature points

The BU-3DFE database was recently developed by Yin et al. at Binghamton University. It was designed to sample 3D facial behaviors with different prototypical emotional states. There are 2500 3D facial expression models in the database which are well distributed across different ethnic origins. Each 3D face model consists of a set of 83 facial feature points, which gives a complete 3D description of a face under a specific facial expression. In this paper, we use all of these 83 facial feature points as shown in Figure 1. α_i, a vector expressing the 3D coordinates of a facial feature point can be described as

$$\alpha_i = (x_i, y_i, z_i), \quad \forall i \in \{1, 2, \cdots, 83\}. \tag{1}$$

The 83 facial feature points on a 3D facial expression model produces $C_{83}^2 = 3403$ unique pairs of facial feature points. The distance, $\delta_{i,j}$, of each pair is normalized by the distance between two outer eye corners, ω, of the same 3D facial expression model in order to make the features scale invariant. The normalized facial feature points are used to form 3D distance vectors, DV_i, for N facial expression models given by the following equations:

$$\delta_{i,j} = \frac{\|\alpha_i - \alpha_j\|}{\omega}, \quad i < j \tag{2}$$

$$DV_i = \begin{pmatrix} \delta_{1,2} \\ \delta_{1,3} \\ \vdots \\ \delta_{2,3} \\ \vdots \\ \delta_{82,83} \end{pmatrix}_{d\times 1}, \forall i \in \{1, 2, \cdots, N\}, \quad d = 3403. \tag{3}$$

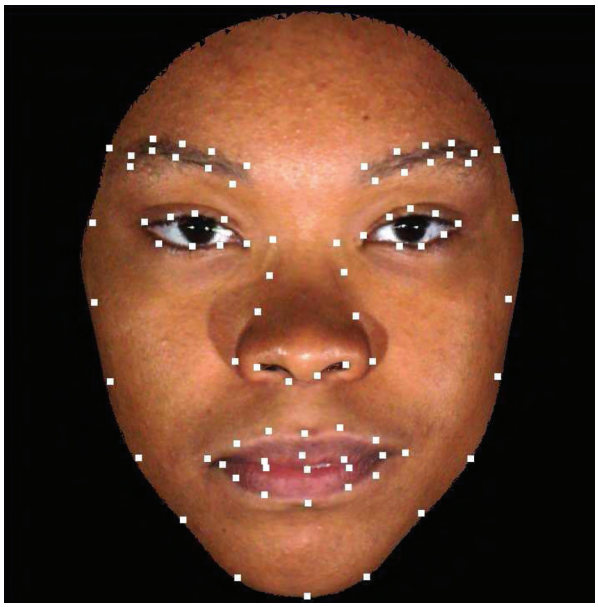

Figure 1. 3D face model consists of a set of 83 facial feature points.

2.2. Construction of an optimal subspace

We propose a method to construct an optimal projection subspace for 3D facial expression recognition. PCA and LDA play a critical role in many pattern classification tasks. PCA is an unsupervised linear feature extraction method that generates a set of orthogonal basis vectors, which describes major variations in the whole training set. PCA seeks the linear transformation matrix W_{PCA} that maps the original space onto an m-dimensional subspace, with $m << d$.

Considering a learning set containing different class samples, we first perform a dimensionality reduction by applying PCA. We then search for the most discriminant projection along eigenvectors by successively selecting the principal components, pc_{best}, in the order of their importance for the recognition of the facial expressions.

We consider a training set of vectors, distributed into c classes. Each vector is then projected into an eigenspace, spanned by m eigenvectors. The selection method consists in seeking, among m principal components, k components $k < m$ which are most discriminant for the 3D facial expression recognition.

We use an iterative process that successively selects components step by step to construct optimal components. The selection criterion Φ is used to define the optimality of a set of components as a general class separability measure, defined by the Fisher criterion, which is expressed as

$$\Phi = \frac{|S_B|}{|S_W|}, \tag{4}$$

where $|S_W|$ and $|S_B|$are, respectively, the determinant of the within-class and between-class scatter matrices. Let y_i^j denote an $m-$dimensional feature vector, extracted from the i^{th} projected sample of the j^{th} class c_j composed of N_j samples. Let μ_j $(j = 1, \cdots c)$ be the mean vector of j^{th} class and μ be the total mean vector in this m-dimensional projection feature space. The within-class and between-class scatter matrices can be

calculated in this feature space as

$$S_W = \sum_{j=1}^{c} \sum_{i=1}^{N_j} \left(y_i^j - \mu_j\right)^T \left(y_i^j - \mu_j\right) \tag{5}$$

$$S_B = \sum_{j=1}^{c} (\mu_j - \mu)^T (\mu_j - \mu). \tag{6}$$

Φ has to be maximized in order to select the best discriminant principal component, pc_{best}. In order to avoid over-fitting and achieve better generalization performances, the selection criterion is computed as the average of Φ over N_{iter} randomly selected learning sets sampled from the original data set. Hence, we use Φ to select the optimal set of components. The classification error rate could have been used for such a selection, but Φ seems to exhibit more stability than the classification error rate, especially when the size and the number of validation sets are small. It should be noted that if the number of features selected is too small compared to the dimensionality of the samples, S_W and S_B are very close to being singular. Consequently, Φ may lead to undesired results. For that reason, as suggested by Labay et al. [9], we have estimated S_W and S_B by using singular value decomposition (SVD). Since S_W and S_B are symmetric and nonnegative definite, product of $k-1$ singular values is used to approximate the determinants of S_W and S_B.

Finally, LDA is computed into the optimum subspace to generate a $(c-1)$-dimensional discriminant subspace, where there are only $c-1$ nonzero eigenvalues corresponding the respective eigenvectors. LDA searches for those vectors in the underlying space that best discriminate among classes.

According to equations 4, 5, 6, the Optimal Feature Selection Procedure is given in Algorithm 1. In the algorithm we consider a training set of vectors, distributed into c classes. Each vector is then projected in an eigenspace (computed by PCA), spanned by m eigenvectors. The selection algorithm consists of seeking among the m principal components the k principal components which are most discriminant for the specific recognition problem, which form the "optimal subspace." We use an iterative process that successively selects principal components step by step to construct an optimal subspace: during steps $\{j = 1 \text{ to } k\}$ we seek the component, among the $\{pc = 1 \text{ to } (k-j+1)\}$ available, which, when added to those previously selected, forms an optimal set of components.

For each iteration we randomly choose a learning set (according to Table 1). The tested principal component, $V(pc)$, is added to those previously kept to build the eigenspace, W_{PCA} with the corresponding eigenvectors.

Table 1. System simulation parameters.

m	k	N_{iter}	*Data*	*learning_set*	*test_set*
20	7	50	420 Subjects	336 Subjects	84 Subjects

Fisher criterion value, $\phi(pc, iter)$, which uses S_W and S_B is calculated for each learning set that is projected into the eigenspace, $Pr\_learning$. The principal component, pc_{best}, with maximum average Fisher criterion over the N_{iter} iterations is added to those previously kept to form the j-dimensional subspace PCA_{optimum}. Finally, LDA is computed into this eigenspace, PCA_{optimum}, to generate $(c-1)$-dimensional discriminant subspace, $\text{Subspace}_{\text{optimum}}$.

Algorithm 1. Optimal feature selection procedure.

$V \leftarrow component\_set(Data, m)$
$PCA_{optimum} \leftarrow \emptyset$
for $j = 1$ *to* k **do**
  $F \leftarrow \emptyset$
  for $iter = 1$ *to* N_{iter} **do**
    $learning\_set \leftarrow Random\_Data(Data)$
    for $pc = 1$ *to* $(k - j + 1)$ **do**
      $W_{PCA} \leftarrow PCA_{optimum} \bigcup V(pc)$
      $Pr\_learning \leftarrow W_{PCA}^{T} * learning\_set$
      $[S_B, S_W] \leftarrow compute\_Fisher(Pr\_learning)$
      $\phi(pc, iter) \leftarrow \frac{|S_B|}{|S_W|}$
      $F(pc) = F(pc) + \phi(pc, iter)$
    end for
  end for
  $pc_{best} \leftarrow \arg\max\{\frac{F(pc)}{N_{iter}}, \forall pc\}$
  $PCA_{optimum} \leftarrow PCA_{optimum} \bigcup pc_{best}$
end for
$Subspace_{optimum} \leftarrow LDA(Data, PCA_{optimum})$

3. Coarse-to-fine classification process

We have tested our PNN setup on the BU-3DFE database, which contains facial expression images with seven fundamental emotional states ($c = 7$): Neutral, Anger, Disgust, Fear, Happiness, Sadness, and Surprise (see Figure 2).

Figure 2. Seven facial expression images for fundamental emotional states.

The simulation parameters for the optimization procedure are given in Table 1. We propose a classification process, using a decision tree-based classifier that takes into account the properties of our representation subspace. This classifier is trained by an iterative selection of individual features that are more salient at each

node of the tree. The fundamental problem when constructing a decision tree is to determine tree partitions based on the training data. Table 2 shows, the confusion matrix of the PNN classifier based facial expression recognition, which contains average recognition rates for each expression. The PNN is based the $(c-1)$-dimensional discriminant subspace.

Table 2. Average confusion matrix showing facial expression recognition rates (%).

%	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise
Neutral	88.93%	2.80%	0.00%	1.55%	0.00%	6.73%	0.00%
Anger	4.64%	85.24%	2.50%	1.19%	0.00%	6.43%	0.00%
Disgust	0.00%	3.33%	87.62%	2.68%	3.04%	0.18%	3.15%
Fear	3.10%	1.55%	3.87%	84.76%	4.05%	1.61%	1.07%
Happy	0.00%	0.00%	1.79%	2.86%	93.93%	0.44%	0.95%
Sadness	8.27%	6.43%	0.00%	2.20%	0.30%	82.80%	0.00%
Surprise	0.00%	0.00%	3.39%	0.18%	0.36%	0.00%	96.07%

The 2D-projections of the seven-facial expression data set along the first two axes of the basis of the Fisherspace generated through the LDA process are illustrated in Figure 3. Experimentally, we observe that the seven-facial expression classes can be regrouped into three main clusters: Group 1 (G1: Surprise), Group 2 (G2: Anger, Sadness and Neutral), Group 3 (G3: Happy, Disgust and Fear).

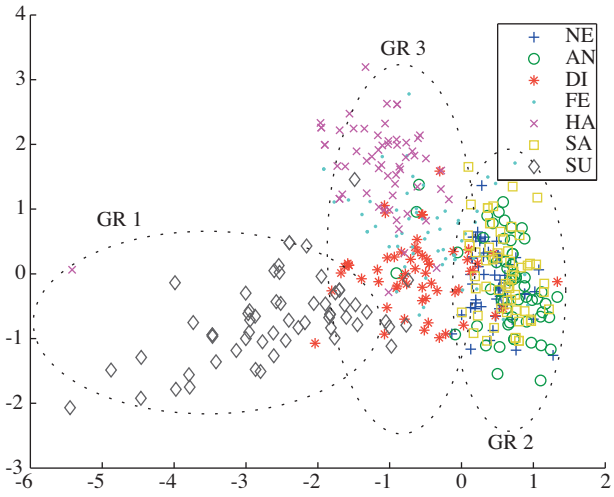

Figure 3. Dataset projected along the first two components of the optimal basis of the Fisherspace generated through the LDA process.

The tree classification is carried out by successive projections: a new sample is first projected onto a coarse representation subspace, where it is associated with the closest class group of facial expression. The sample is then projected onto a finer representation subspace, describing the classes belonging to the group, to recognize its expression. The classification is based on the Euclidean distance. The coarse representation subspace of the main clusters is shown in Figure 4. A new sample is first projected onto G. If its projection is closer to Group 1, it is classified as Surprise. If it is closer to Group 2, it is projected onto a finer representation subspace S_{G2} (Figure 5) and then classified into the nearest facial expression class (Anger, Sadness or Neutral).

If it is closer to Group 3, it is projected onto a finer representation subspace S_{G3} (Figure 6), then classified into the nearest facial expression class (Disgust, Fear or Happy). The procedure applied to classify a new sample x is described in Algorithm 2.

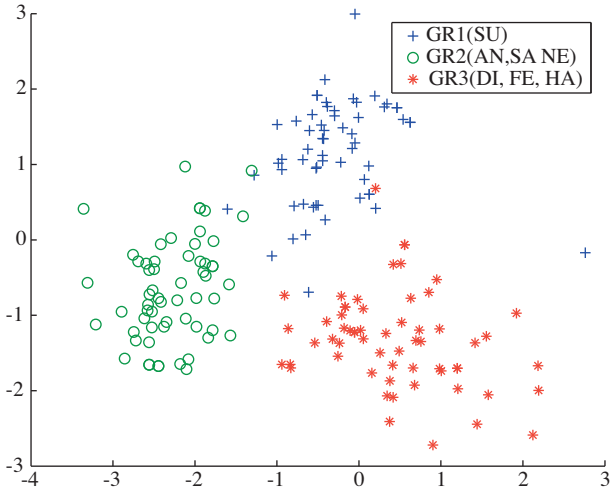

Figure 4. The coarse representation subspace G of the main clusters.

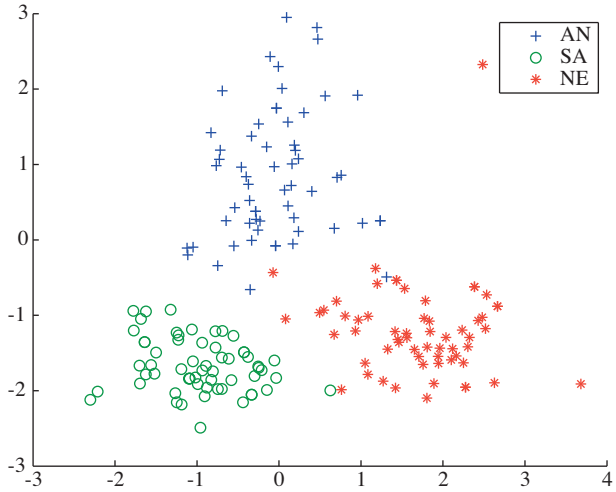

Figure 5. Group 2 projected along the first two components of the finer representation subspace S_{G2}.

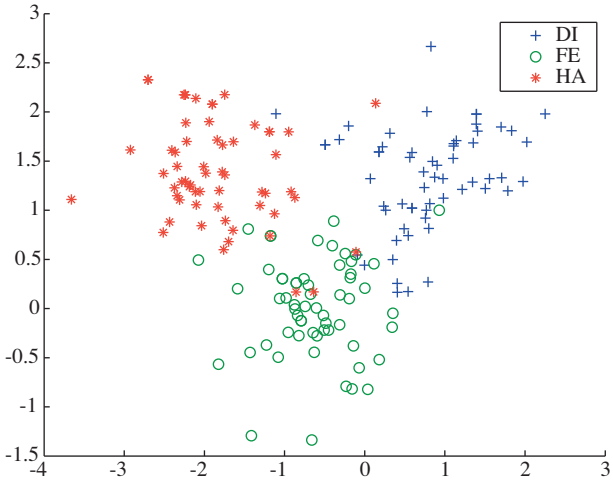

Figure 6. Group 3 projected along the first two components of the finer representation subspace S_{G3}.

4. Results and conclusions

In this paper, we have proposed a statistical technique to construct optimal subspace for 3D facial expression recognition. The proposed technique uses fisher criterion based feature selection process in order to select the optimal feature vectors generated from facial expression vectors extracted by using distances of 3D facial feature points. The optimally selected features are used with PNN classifier for facial expression recognition with an average performance of 88.5%. The addition of coarse-to-fine approach, in the form of tree-PNN, has

increased the overall facial expression recognition average performance to 93.7%. Compared to the existing image based approaches [3, 6, 7] and model based approach cite [8, 10, 11], our optimized feature selection based approach shows superior performance as a result of the illumination and orientation invariance achieved by using 3D geometrically localized facial feature points. The results illustrated in Table 3 show that choosing an optimal representation for facial expressions using fisher criterion improves the performance of the facial expression recognition process. Additional performance improvement is achieved by using tree-PNN within the coarse-to-fine approach for the classification.

Algorithm 2. Tree-PNN based classifier procedure.

1: *define y as the projection x onto G*
2: **if** $\{y \in G_1\}$**then**
3: $class \leftarrow Surprise$
4: **else if** $\{y \in G_2\}$**then**
5: *define y as the projection x onto* G_2
6: *classify z into the nearest class*
7: $class \leftarrow \{Anger,\ Sadness,\ Neutral\}$
8: **else** $\{y \in G_3\}$
9: *define y as the projection x onto* G_3
10: *classify z into the nearest class*
11: $class \leftarrow \{Disgust,\ Fear,\ Happy\}$
12: **end if**

Table 3. Performance comparison.

Method	Neutral	Anger	Disgust	Fear	Happy	Sadness	Surprise	Average
Wang et al. (LDA) [5]		80.0%	80.4%	75.0%	95.0%	80.4%	90.8%	83.6%
Tang et al. (SVM) [6]	-	86.7%	84.2%	74.2%	95.8%	82.5%	99.2%	87.1%
Tang et al. (NBC) [7]	-	91.7%	90.0%	75.8%	90.8%	80.0%	97.5%	87.6%
Mpiperis et al. (NBC) [8]	-	83.6%	100.0%	97.9%	99.2%	62.4%	100.0%	90.5%
Soyel et al. (FFNN) [10]	86.7%	85.0%	91.7%	91.7%	95.0%	90.7%	98.3%	91.3%
Soyel et al. (PNN) [11]	84.8%	83.3%	85.1%	82.6%	95.3%	86.4%	97.7%	87.8%
Prop. Method (Tree-PNN)	96.1%	91.7%	93.9%	90.6%	94.1%	90.8%	98.9%	93.7%

References

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

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

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

[4] L. Yin, X. Z. Wei, M. Rosato, "A 3D facial expression database for facial behavior research", 7th International Conference on Automatic Face and Gesture Recognition, pp. 211-216, 2006.

[5] J. Wang, L. Yin, X. Wei, Y. Sun, "3D facial expression recognition based on primitive surface feature distribution", Computer Vision and Pattern Recognition, Vol. 2, pp. 1399 -1406, 2006.

[6] H. Tang, T. S. Huang, "3D Facial expression recognition based on properties of line segments connecting facial feature points", IEEE International Conference on Automatic Face and Gesture Recognition, pp.1-6, 2008.

[7] H. Tang, T. S. Huang, "3D facial expression recognition based on automatically selected features", Computer Vision and Pattern Recognition, pp.1-8, 2008.

[8] Mpiperis, S. Malassiotis, M. G. Strintzis, "Bilinear Models for 3D Face and Facial Expression Recognition", IEEE Transactions on Information Forensics and Security, Vol. 3, pp. 498-511, 2008.

[9] V. A. Labay, J. Bornemann, "Matrix Singular Value Decomposition for Pole-Free Solutions of Homogeneous Matrix Equations as Applied to Numerical Modeling Methods", IEEE Microwave and Guided Wave Letters, Vol.2, pp.49-51, 1992.

[10] H. Soyel, H. Demirel, "Facial Expression Recognition using 3D Facial Feature Distances", Lecture Notes in Computer Science, Vol. 4633, pp. 831-838, 2007.

[11] H. Soyel, H. Demirel, "3D Facial Expression Recognition with Geometrically Localized Facial Features", Computer and Information Sciences, pp. 1-4, 2008.