Facial Expression Recognition with Emotion-Based Feature Fusion

Cigdem Turan[1], Kin-Man Lam[1], Xiangjian He[2]
[1]The Hong Kong Polytechnic University, Hong Kong, SAR, [2]University of Technology Sydney, Australia
E-mail: cigdem.turan@connect.polyu.hk, enkmlam@polyu.edu.hk, Xiangjian.He@uts.edu.au

Abstract— **In this paper, we propose an emotion-based feature fusion method using the Discriminant-Analysis of Canonical Correlations (DCC) for facial expression recognition. There have been many image features or descriptors proposed for facial expression recognition. For the different features, they may be more accurate for the recognition of different expressions. In our proposed method, four effective descriptors for facial expression representation, namely Local Binary Pattern (LBP), Local Phase Quantization (LPQ), Weber Local Descriptor (WLD), and Pyramid of Histogram of Oriented Gradients (PHOG), are considered. Supervised Locality Preserving Projection (SLPP) is applied to the respective features for dimensionality reduction and manifold learning. Experiments show that descriptors are also sensitive to the conditions of images, such as race, lighting, pose, etc. Thus, an adaptive descriptor selection algorithm is proposed, which determines the best two features for each expression class on a given training set. These two features are fused, so as to achieve a higher recognition rate for each expression. In our experiments, the JAFFE and BAUM-2 databases are used, and experiment results show that the descriptor selection step increases the recognition rate up to 2%.**

I. INTRODUCTION

Facial expression recognition (FER) is one of the most interesting topics in the field of human-computer interaction, and has become a popular research topic during the last few decades. Before training classifiers for recognizing facial expressions, feature extraction is performed from face images in order to extract the distinctive features which can distinguish the different expressions.

The features used for FER can be divided into two categories: geometric-based and appearance-based methods. Geometric-based features benefit from the shape and location information of facial components such as the eyes, mouth and eyebrows, while appearance-based features contain changes in the skin texture such as wrinkles, bulges and furrows. To a certain extent, these two types of features are supplementary to each other.

In this paper, four competent local descriptors are selected, and their performances for facial expression recognition are evaluated. These four descriptors are Local Binary Pattern (LBP) [1], Local Phase Quantization (LPQ) [2], Weber Local Descriptor (WLD) [3], and Pyramid of Histogram or Oriented Gradients (PHOG) [4], which have been used for facial expression recognition in the literature [5-8].

It can be seen that, from confusion matrices, different descriptors can achieve different recognition rates for a specific emotion. In the past, a single local descriptor was usually studied to achieve the best overall performance for all emotions. In this paper, we propose to identify the best two features for each expression, which are then fused to form a coherent feature for representing a particular expression.

Manifold learning aims to embed high-dimensional data in a lower dimensional space while preserving the intrinsic characteristics. In [9], Shan et al. compared the performances of different manifold learning techniques on facial expression recognition, and showed that Supervised Locality Preserving Projections (SLPP) [10] achieves the best performance. More importantly, SLPP also considers the class information in the construction of the manifolds.

According to [11], emotions can be classified into four basic classes: 1) Anger-Disgust (AN-DI), 2) Fear-Surprise (Fe-SU), 3) Sadness (SA), and 4) Happiness (HA). In a video sequence, the set of specific facial movements of a particular emotion does not occur at once but sequentially over time. In the early stages of anger or disgust, accurate discrimination between these two expressions is not obvious, similar to that between fear and surprise. Based on this, the number of expression classes is set at four, and the performances of the respective feature descriptors are measured for each of the expression classes. Then, the best two descriptors for each expression are identified and fused using Discriminant-Analysis of Canonical Correlations (DCC) [12] to form a coherent feature set. Our aim is to find the best discriminant features by combining the different descriptors for recognizing each facial expression. To the best of our knowledge, we are the first to use different coherent descriptors for the recognition of different expressions. Based on the coherent features, a classifier is learned for each expression. In other words, four classifiers are learned for the four expressions, i.e. anger-disgust, fear-surprise, happiness, and sadness.

The rest of this paper is organized as follows: The details of our proposed approach are presented in Section II. In Section III, experimental setup is described, and the experimental results are shown. Section IV concludes the paper.

II. DETAILS OF OUR APPROACH

Before extracting features, the faces are scaled and aligned based on the position of the eyes such that the distance between the two eyes is 64 pixels and the image size is 126×100 pixels. In order to obtain more effective facial

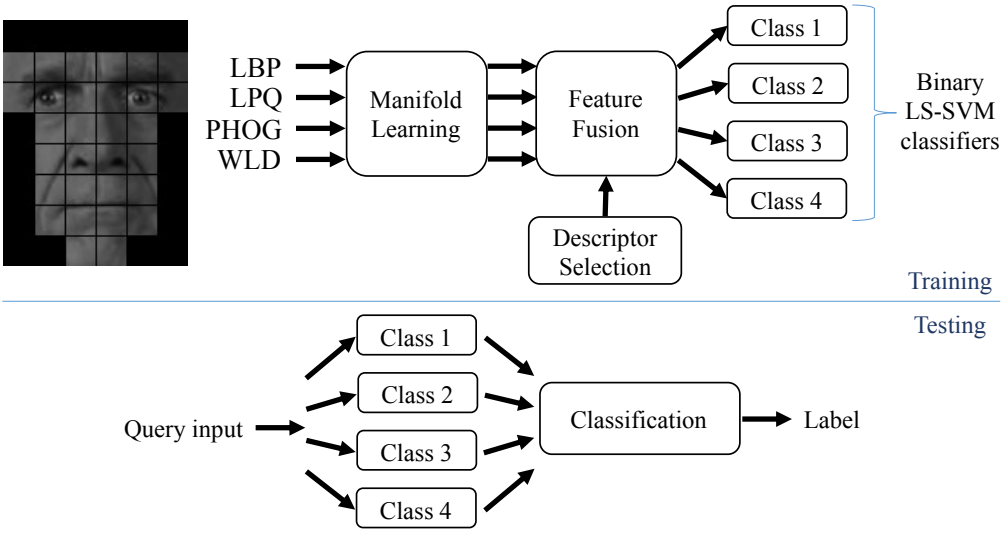


Fig. 1 The emotion-based feature fusion scheme for facial expression recognition.

features, each image is divided into 8×6 regions, and 30 of the regions are used for feature extraction, as illustrated in Fig. 1. We can see that the selected regions contain the salient facial features, so they can represent facial expressions more effectively. After extracting the features, i.e. LBP, LPQ, PHOG, and WLD, supervised LPP is applied for manifold learning.

In the rest of this section, first, the four descriptors, SLPP, and DCC, are explained in detail. Then, the process of evaluating the performance of each descriptor for each expression class is described. Finally, the proposed adaptive descriptor selection algorithm is presented.

A. The Local Descriptors

In this paper, four different local descriptors are considered, because: 1) they have been used widely for facial expression recognition, and 2) they represent facial expressions in terms of different aspects such as intensity, phase, and shape.

The first descriptor used in our approach is Local Binary Pattern (LBP) [13], which was proposed as a texture descriptor. In LBP, the label for each pixel is represented as an 8-bit binary number by thresholding the 3×3 neighboring pixels with the center pixel value. The feature vector for the considering region is then represented using a 256-bin histogram. The advantage of LBP is that it is insensitive to monotonic variations caused by illumination changes.

The second descriptor considered is Local Phase Quantization (LPQ) [2], which was also proposed as a texture descriptor. Unlike LBP, which uses intensity value, LPQ is based on the blur invariance property of the Fourier phase information with the assumption that the blur is centrally symmetric. LPQ computes the short-term Fourier transform at each pixel over a rectangular $M \times M$ neighborhood. Using the local Fourier coefficients at four different frequencies, the phase information is recovered by using a scalar quantizer resulting in an 8-bit number, represented as a decimal number between 0 and 255. The distribution of the numbers is then represented using a histogram.

An extension of the Histogram of Oriented Gradients (HOG) [14] descriptor, the Pyramid of Histogram of Oriented Gradients (PHOG) [4], is a descriptor commonly used for object recognition. PHOG represents an image using its local shape at different scales. The Canny edge detector is applied to an image, which is then divided into spatial cells based on the number of levels. At each pyramid level, the orientation gradients of the edge contours are calculated using the 3×3 Sobel masks. The orientation gradients are represented by using a K-bin histogram followed by concatenation of the histograms of each level. The final feature vector is of dimension $K \times \Sigma 4l$, where l is the number of pyramid levels and K is the number of bins in the histograms. In our experiments, l and K are set at 2 and 8, respectively.

Weber Local Descriptor (WLD), proposed by Chen et al. [3], is an image descriptor which is derived from the Weber's Law, which states that human perception of change in a given stimulus also depends on the intensity of the original stimulus. According to this law, the change of a stimulus can be recognized if the ratio of the change to the original stimulus is larger than a certain value. WLD consists of two components: differential excitation and orientation. Differential excitation considers the ratio between the current pixel and the relative intensity differences against it. The second component, i.e. orientation, is the ratio between the vertical and horizontal gradients. Weber magnitude δ_m and orientation δ_o are defined as follows:

$$\delta_m(x_c) = \cos^{-1}\left(\alpha \sum_{i=0}^{p-1} \frac{x_i - x_c}{x_c}\right), \quad (1)$$

$$\delta_o(x_c) = \cos^{-1}\frac{x_1 - x_5}{x_3 - x_7}, \quad (2)$$

x_0	x_7	x_6
x_1	x_c	x_5
x_2	x_3	x_4

Fig. 2 The mask used for WLD.

where x_c denotes the center pixel and x_i is the neighboring pixels, where $i = 0, 1, \ldots, p-1$, as illustrated in Fig. 2. Weber magnitude δ_m and orientation δ_o are quantized and represented by using a 2D histogram. This histogram is then mapped to a 1D histogram to obtain the feature vector.

B. Supervised Locality Preserving Projection (SLPP)

Locality Preserving Projection (LPP) [15], which is a linear approximation of the nonlinear Laplacian Eigenmap [16], employs the following minimization problem:

$$\min_{w} \sum_{i,j} \left(w^T x_i - w^T x_j\right)^2 s_{ij},$$

where **S** = [s_{ij}] is the similarity matrix that preserves the local neighborhood information. An edge is added between nodes *i* and *j* if x_i and x_j are among the *k* nearest neighbors of each other. Heat kernel sets the edge weight s_{ij} as above if there is an edge between nodes *i* and *j*:

$$s_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}},$$

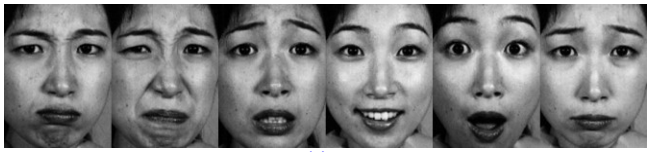
where *t* is the parameter for the method. An extension of LPP, namely supervised LPP [10], uses the class information when constructing the similarity matrix. In other words, an edge is added between nodes *i* and *j* if and only if x_i and x_j belong the same class and are among the *k* nearest neighbors of each other.

TABLE I
EXPERIMENT RESULTS FOR JAFFE DATABASE

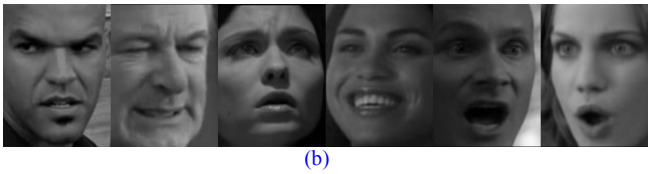
JAFFE	LBP	LPQ	WLD	PHOG
AN-DI	**90.49% ± 1.84%**	85.79% ± 2.97%	89.51% ± 0.90%	**95.85% ± 0.49%**
FE-SU	93.11% ± 1.20%	86.34% ± 1.02%	**96.07% ± 1.05%**	**95.85% ± 1.37%**
HA	**96.28% ± 1.30%**	92.35% ± 1.02%	96.07% ± 1.52%	**97.27% ± 0.39%**
SA	**91.04% ± 1.48%**	88.09% ± 0.81%	**90.82% ± 1.12%**	89.29% ± 0.73%
ALL	**89.18% ± 0.46%**	83.28% ± 0.62%	89.18% ± 0.71%	**90.60% ± 0.71%**

TABLE II
EXPERIMENT RESULTS FOR BAUM-2 DATASET

BAUM-2	LBP	LPQ	WLD	PHOG
AN-DI	73.92% ± 0.93%	**76.96% ± 0.28%**	**74.98% ± 0.37%**	70.16% ± 0.35%
FE-SU	81.91% ± 0.51%	**83.28% ± 0.56%**	**82.36% ± 0.65%**	81.54% ± 0.12%
HA	**88.37% ± 0.54%**	**89.82% ± 0.46%**	88.25% ± 0.25%	87.48% ± 0.36%
SA	**85.48% ± 0.31%**	**85.62% ± 0.45%**	84.54% ± 0.62%	84.70% ± 0.42%
ALL	62.85% ± 0.62%	**66.71% ± 0.56%**	**63.35% ± 0.48%**	59.86% ± 0.39%



(a)



(b)

Fig. 3 Sample images for (a) the JAFFE, and (b) the BAUM-2 databases.

C. Discriminant-Analysis of Canonical Correlations (DCC)

Discriminant-Analysis of Canonical Correlations (DCC) [12] was proposed as a discriminative learning method by Kim et al., inspired by Linear Discriminant Analysis (LDA) [17] which has been used commonly for dimension reduction aiming to preserve the class discriminatory information. Similar to LDA, DCC seeks to find a transformation matrix *W* for two feature sets *X* and *Y* such that $X_w = W^T X$ and $Y_w = W^T Y$, where the matrix *W* maximizes the canonical correlations of the within-class sets, while minimizing the canonical correlations of the between-class sets.

In this paper, DCC is applied to two different feature sets extracted using two different descriptors in order to fuse them in a manner that the transformed feature set will have the

TABLE III
EXPERIMENT RESULTS FOR BAUM-2 + JAFFE DATABASE

BAUM-2 + JAFFE	LBP	LPQ	WLD	PHOG
AN-DI	74.19% ± 0.28%	**77.35% ± 0.27%**	74.68% ± 0.77%	**74.92% ± 0.11%**
FE-SU	82.21% ± 0.25%	82.59% ± 0.08%	**83.24% ± 0.19%**	**83.26% ± 0.56%**
HA	**89.92% ± 0.28%**	**90.08% ± 0.20%**	88.99% ± 0.27%	88.62% ± 0.26%
SA	84.55% ± 0.28%	**86.50% ± 0.32%**	**84.84% ± 0.85%**	84.66% ± 0.18%
ALL	64.74% ± 0.35%	**68.85% ± 0.37%**	**66.48% ± 0.52%**	64.88% ± 0.46%

most discriminant, coherent features to represent each emotion class.

D. Evaluating the Descriptors

In this paper, we evaluate the performance of each descriptor using the one-versus-all classification scheme. The features of those face images of a particular emotion are labeled as positive, while those of other emotions as negative. Then, a binary classifier is trained using Support Vector Machine (SVM) for each class of emotion, so there are a total of 4 classifiers. The recognition rate for each of the descriptors is measured. In addition, the two best descriptors for each emotion class are paired and then fused using DCC to form a single coherent descriptor. The performances of these coherent features are also evaluated using the one-versus-all scheme.

E. The Proposed Automatic Descriptor Selection Algorithm

In the evaluation of the respective descriptors and the coherent descriptors, we found that fusing the two descriptors which achieve the highest recognition rates for a particular emotion can achieve higher accuracy than the individual descriptors. However, the best descriptors for each emotion may be different, as well as for different databases. Thus, fusing fixed descriptors to form a coherent descriptor is not the optimum way to achieve the best results. To achieve robust facial expression recognition, an adaptive descriptor selection step is included in our algorithm. The descriptor selection algorithm analyzes the performances of each pair of descriptors for each expression class on the given training set and determines the best two descriptors for each expression class regarding the training set; a total of 4 pairs of descriptors are selected. As observed before, the best two descriptors may be different for different expression classes. Therefore, a pair of best descriptors is determined for each expression class. In the descriptor selection step, *N*-fold cross validation, where $N = 3$ in our experiments, has been conducted on the training set. After identifying the best descriptors, a binary classifier is trained for each class using the most salient features, which are created by fusing the two best features by using DCC. For a query input, four different feature vectors are created and tested on the four different classifiers. The output of each of the classifiers is viewed as the probability of the query belonging to the corresponding class. The query is assigned to the class whose corresponding output has the highest value.

III. Experimental Protocol and Results

A. Experimental Protocol

Experiments were conducted on three databases: BAUM-2, JAFFE, and a combination of two databases. JAFFE [18, 19] consists of images from 10 Japanese females that express 6 basic emotions and the neutral. Unlike JAFFE which is a database recorded in a controlled environment, the BAUM-2 [20] database consists of expression videos, extracted from movies. In our experiments, an image dataset, namely BAUM-2i, consisting of images with peak expressions from the videos from BAUM-2 is considered. There are 183 face images from 10 subjects in the JAFFE database that express 6 basic emotions, while there are 829 face images from 250 subjects in the BAUM-2i static expression dataset. Since the BAUM-2 database was created by extracting from movies, the images are in the close-to-real-life conditions (i.e. with pose, age, and illumination variations, etc.) and are more challenging than those in an acted database, as seen in Fig. 3.

TABLE IV
COMPARISON OF THE PERFORMANCES OF BEST DESCRIPTORS OF EACH DATASET WITH ADAPTIVE DESCRIPTOR SELECTION METHOD

	JAFFE	BAUM-2i	BAUM-2i + JAFFE
LBP-PHOG	91.58% ± 0.30%	67.00% ± 0.50%	69.23% ± 0.15%
LPQ-WLD	87.32% ± 0.46%	68.47% ± 0.41%	69.96% ± 0.60%
Adaptive Descriptor Selection	**92.13% ± 0.91%**	**68.71% ± 0.53%**	**70.99% ± 1.13%**

It has been shown that SVM can achieve satisfactory results even for high-dimensional feature vectors. Furthermore, the more recent Least Square SVM (LS-SVM) [21] has been proposed, which is very efficient on large datasets since it uses linear programming, rather than convex programming in SVM. LS-SVM has been applied to different recognition problems like face [22] and facial expression [23, 24]. Therefore, our proposed method uses LS-SVM [25] with the Gaussian kernel.

B. Experiment Results for the Evaluation of the Descriptors

To evaluate the performances of the selected descriptors, 5-fold cross validation was used. In this experiment, it is aimed to present that the performance of each descriptor is different for the different expression classes. Table I shows the performances of the different descriptors based on the JAFFE dataset. PHOG can achieve the highest accuracy for the expression classes Anger-Disgust and Happiness, while WLD performs better for the class Fear-Surprise. LBP descriptor outperforms other descriptors for the class Sadness. The overall performances of each of the descriptors for all the expression classes are also evaluated. As observed, the overall performances of the classifiers are less than the performances of any other binary classifiers. The reason behind it is that the overall performance considers all the four labels, while the binary classifiers consider the labels as positive and negative. From the results, we can see that PHOG and LBP are the two best descriptors for recognizing all the expressions. Similarly, Table II shows the corresponding performances based on the BAUM-2i dataset. LPQ outperforms all other descriptors for all the expression classes. LPQ and WLD achieve the best overall performances.

As observed, even for the same expression classes, different descriptors can achieve the best recognition rates with different datasets. The reason for this is due to the fact that the two datasets are different in terms of race, age, resolution, pose, etc. Thus, the two databases are also merged

into a single one to explore the form a database with images having more variations. The two best descriptors are then identified for each expression class. Table III shows the performances of the descriptors with respect to each of the expression classes. It can be seen that the two best descriptors selected based on BAUM-2 + JAFFE are correlated with the two best descriptors of either dataset. For instance, LPQ and PHOG descriptors achieve the highest accuracies for the AN-DI expression class in BAUM-2 + JAFFE (first row of the results in Table III). We can also observe that LPQ and PHOG are the descriptors that can achieve the best performances for the AN-DI class on BAUM-2 and JAFFE, respectively.

The results, once again, show that the different expression classes of different datasets can be represented more effectively by a different set of descriptors. Thus, the descriptors to be used for classification should not be fixed for a specific expression class, and should be adaptive to the expressions and the image conditions.

C. *Experiment Results for the Proposed Adaptive Descriptor Selection Algorithm*

Based on the results in Tables I, II, and III, the descriptors to be used are adaptive to the expression classes. For the JAFFE database, the fused features for the AN-DI, FE-SU, HA, and SA are PHOG+LBP, WLD+PHOG, PHOG+LBP, and LBP+WLD, respectively. For the BAUM-2i database, the fused features for the AN-DI, FE-SU, HA, and SA are LPQ+WLD, LPQ+WLD, LPQ+LBP, and LPQ+LBP, respectively. For the combined database, i.e. BAUM-2i + JAFFE, the fused features for the AN-DI, FE-SU, HA, and SA are LPQ+PHOG, WLD+PHOG, LPQ+LBP, and LPQ+WLD, respectively. We compare our proposed adaptive algorithm with the non-adaptive algorithm, which uses the same fused features for all the expression classes. For the JAFFE and BAUM-2i databases, PHOG+LBP and LPQ+WLD, respectively, achieve the best overall performance. These two fused features are used non-adaptively for the recognition of all the expression classes. In the experiments, 5-fold cross-validation has been conducted.

As shown in Table IV, using fused features can achieve higher recognition rates than the individual descriptors, and the adaptive algorithm outperforms the non-adaptive one. Also, as observed, the adaptive descriptor selection algorithm increases the accuracy up to 2% for the JAFFE, BAUM-2i and BAUM-2 + JAFFE datasets since the most salient features are used in the recognition of each expression class. The recognition rate for the BAUM-2i dataset is lower than that for JAFFE since BAUM-2i was created with expression images extracted from movies. This makes the dataset more challenging because of the pose, illumination and resolution variations.

IV. CONCLUSIONS

In this paper, we aim to show the differences in the performances regarding four commonly used descriptors: LBP, LPQ, PHOG and WLD. SLPP is applied as the manifold learning method, which preserves the locality information with the help of class information. Then, DCC is adopted to fuse the best two feature sets by projecting them into a coherent subspace. We have proposed a classification method, which utilizes the adaptive descriptor selection algorithm to further increase the performance of a facial expression recognition system. In our experiments, four expression classes are considered for evaluating the performance of the proposed classification method. The LS-SVM is employed based on the features projected to a coherent subspace to learn a binary classifier for each of the expression classes. Experiment results have shown that the proposed classification method can achieve higher recognition rate than any of the individual descriptors.

REFERENCES

[1] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. 24(7): p. 971-987.

[2] V. Ojansivu and J. Heikkilä, Blur insensitive texture classification using local phase quantization, in Image and signal processing, 2008, Springer. p. 236-243.

[3] J. Chen et al., WLD: A robust local image descriptor, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. 32(9): p. 1705-1720.

[4] A. Bosch, A. Zisserman, and X. Munoz, Representing shape with a spatial pyramid kernel, in Proceedings of the 6th ACM international conference on Image and video retrieval, 2007, ACM.

[5] A. Dhall et al. Emotion recognition using PHOG and LPQ features, in Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on. 2011. IEEE.

[6] C. Shan, S. Gong, and P.W. McOwan, Facial expression recognition based on local binary patterns: A comprehensive study, Image and Vision Computing, 2009. 27(6): p. 803-816.

[7] S. Liu, Y. Zhang, and K. Liu. Facial expression recognition under partial occlusion based on Weber Local Descriptor histogram and decision fusion, in Control Conference (CCC), 2014 33rd Chinese. 2014, IEEE.

[8] S. Yang and B. Bhanu. Facial expression recognition using emotion avatar image, in Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, 2011, IEEE.

[9] C. Shan, S. Gong, and P.W. McOwan, A comprehensive empirical study on linear subspace methods for facial expression analysis. in Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on, 2006, IEEE.

[10] C. Shan, S. Gong, and P.W. McOwan, Appearance manifold of facial expression, in Computer Vision in Human-Computer Interaction2005, Springer, p. 221-230.

[11] R.E. Jack, O.G. Garrod, and P.G. Schyns, Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time, Current biology, 2014, 24(2): p. 187-192.

[12] T.-K. Kim, J. Kittler, and R. Cipolla, Discriminative learning and recognition of image set classes using canonical correlations, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2007, 29(6): p. 1005-1018.

[13] T. Ojala, M. Pietikäinen, and D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern recognition, 1996, 29(1): p. 51-59.
[14] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in Computer Vision and Pattern Recognition, 2005, CVPR 2005, IEEE Computer Society Conference on, 2005, IEEE.
[15] X. Niyogi, Locality preserving projections. in Neural information processing systems, 2004, MIT.
[16] M. Belkin and P. Niyogi, Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering, in NIPS, 2001.
[17] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman," Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. 19(7): p. 711-720.
[18] M.J. Lyons et al., The Japanese female facial expression (JAFFE) database. 1998.
[19] M.J. Lyons et al. Coding facial expressions with gabor wavelets, in Automatic Face and Gesture Recognition, 1998, Proceedings. Third IEEE International Conference on. 1998. IEEE.
[20] C.E. Erdem, C. Turan, and Z. Aydin, BAUM-2: a multilingual audio-visual affective face database, Multimedia Tools and Applications, 2014: p. 1-31.
[21] J.A. Suykens and J. Vandewalle, Least squares support vector machine classifiers, Neural processing letters, 1999. 9(3): p. 293-300.
[22] J. Xie, Face recognition based on Curvelet transform and LS-SVM. in Proceedings of the 2009 International Symposium on Information Processing (ISIP'09), Huangshan, PR China. 2009. Citeseer.
[23] S. Liu et al. Facial expression recognition approach based on least squares support vector machine with improved particle swarm optimization algorithm, in Robotics and Biomimetics (ROBIO), 2010 IEEE International Conference on. 2010. IEEE.
[24] X. Zhao and S. Zhang, Facial expression recognition based on local binary patterns and least squares support vector machines, in Advances in Electronic Engineering, Communication and Management Vol. 22012, Springer. p. 707-712.
[25] K. De Brabanter et al., LS-SVMlab toolbox user's guide. ESAT-SISTA Technical Report, 2011: p. 10-146.