Facial Expression Identification using Regularized Supervised Distance Preserving Projection

Sohana Jahan, Moriyam Akter, Sifta Yeasmin and Farhana Ahmed Simi

Department of Mathematics, University of Dhaka, Dhaka-1000, Bangladesh

Abstract

Facial expression recognition is one of the most reliable and a key technology of advanced human-computer interaction with the rapid development of computer vision and artificial intelligence. Nowadays, there has been a growing interest in improving expression recognition techniques. In most of the cases, automatic recognition system's efficiency depends on the represented facial expression feature. Even the best classifier may fail to achieve a good recognition rate if inadequate features are provided. Therefore, feature extraction is a crucial step of the facial expression recognition process. In this paper, we have used Regularized Supervised Distance Preserving Projection for extracting the best features of the images. Numerical experiment shows that the use of this technique outperforms many of state of art approaches in terms of recognition rate.

I. Introduction

Facial expression[1,2] is a visible way of human cognitive activity. It is a major way of human emotional communication. It is reported[3] that, facial expression constitutes 55% of the effect of a communicated message while language and voice constitute 7% and 38% respectively. With the rapid development of computer vision and artificial intelligence, facial expression recognition becomes the key technology of advanced human-computer interaction. Nowadays many researchers have been paying attention to expression recognition. In a model-identification system, the recognition rate mainly depends on the extracted features. Feature selection consists in finding the most relevant features from the classification point of view. An efficient selection of main features leads to less error rate. Therefore, feature extraction plays a very important role. Researchers are using various dimension reduction techniques to extract adequate features. Among a great number of dimensionality reduction methods available, the recently proposed *Supervised Distance Preserving Projections*[4] (**SDPP**) method and its regularized version **RSDPP**[5] has shown promising results on different applications[5,6,7,8].

SDPP4 is a supervised method which projects data points from a higher dimensional space $X \subset Rm$ into a lower dimensional space $Z \subset Rr$ with dimensionality $r \ll m$ through the linear function $f : Rm \rightarrow Rr$ defined by $f(x) = W^T x$. The basic formulation of SDPP aims to preserve distances locally between data points in the projected space Z and the output space Y. The method seeks for the transformation matrix W that minimizes the objective function:

$$F(W) = \frac{1}{n}\sum_{i=1}^{n}\sum_{x_j \in N(x_i)}(d_{ij}^2(W) - \delta_{ij}^2)^2$$

where, $N(x_i)$ denotes a neighborhood of x_i, $d_{ij}{}^2(W)$ represents Euclidean metric $d_{ij}{}^2(W) = \|z_i - z_j\|^2$ that is used to characterized the pairwise distances of points in the transformed space and $\delta_{ij}{}^2 = \|y_i - y_j\|^2$ is the distances in output space.

The Regularized Supervised Distance Preserving Projection (RSDPP)[5] is obtained by adding a regularization term $(\lambda vec(W)^T vec(W))$ to the standard SDPP method. RSDPP minimizes the function:

$$J(W) = \frac{1}{n}\sum_{ij} G_{ij}(d_{ij}^2(W) - \delta_{ij}^2)^2 + \lambda vec(W)^T vec(W)$$

The regularization term allows handling ill-conditioning matrix problems. In addition, regularization is especially important when the training sample number is insufficient and distribution is unbalanced seriously. In most of the face data, each image is represented using a high-dimensional vector. Thus, the number of variables (features) is much higher than the number of samples. When classification or recognition tasks are applied, ill-posed problems arise. For such ill-posed problems, regularization improves the recognition rate. RSDPP reduces the dimension of face image data in such a way that projection of the data from higher dimensional space to the lower-dimensional space preserves the intrinsic structure.

The rest of the paper is organized as follows

In the next section, we have discussed the problem formulation of the expression recognition model. The following section includes a brief review of SDPP. In section IV, Regularized SDPP is discussed. Numerical experiments have been done in section V where we have given a comparison of the performance of RSDPP with SDPP and some other state of art approaches like the Eigenface method and Supervised Principal Component Analysis (SPCA) in terms of expression recognition rate. Finally, we have concluded our work in section VI.

II. Problem Formulation

A general expression recognition problem can be stated as follows:

Given a set of face images labelled with the identity of facial expressions (e.g. normal, happy, sad, surprise, anger, disgust,

---
*Author for correspondence. e-mail: sjahan.mat@du.ac.bd

etc.) which we call the training set and an unlabelled set of face images from the same group of people called the testing set. The aim is to identify each expression in the testing set.

In an image recognition problem, an image is considered as a high dimensional vector where each of the coordinates corresponds to a pixel value in the sample image. Working with this huge dimension requires more computation time, large memory space as well as higher computation complexity. So, reduction of the dimension is the first step of the system.
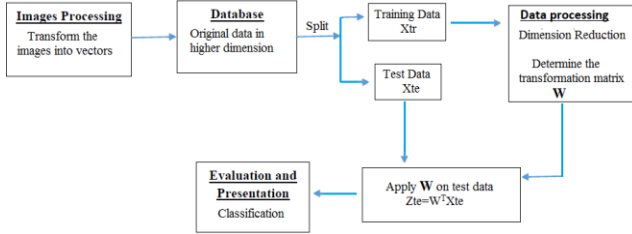

Fig. 1. Overview of Expression Recognition process using dimension reduction.

Given a set of n sample images $x_1, x_2, x_3, \ldots, x_n$ where $x_i \in R^m$. Each image belongs to one of the l classes (l expressions) $C_1, C_2, \ldots, C_l$. Consider a linear transformation $z = W^T x$ from the original m-dimensional image space into a r-dimensional feature space, where $r \ll m$ and $W \in R^{m \times r}$ is the transformation matrix with orthonormal columns.

The problem is to determine the transformation matrix W using the training set (a subset of the given data points) which will be used to select important features to recognize the images in the testing set.

III. Supervised Distance Preserving Projection (SDPP)

The Supervised Distance Preserving Projection (SDPP) is a dimensionality reduction method that minimizes the differences between distances among projected co-variates and distances among responses locally. That means the local geometrical structure of the low dimensional sub-space preserves the geometrical characteristics of the response space. It also preserves the continuity of the response space.

SDPP[4] is proposed recently by Zhu et al. In his paper, the author proposed the following methodology. Suppose X is the higher dimensional space and Y is the response space. Assuming that the mapping $h : X \to Y$ is continuous and provided X is well sampled. It is assumed that for each point $x \in X$ and for every $\varepsilon_y > 0$ there exists an $\varepsilon_x > 0$ such that $d(x, x') < \varepsilon_x \Rightarrow \delta(h(x), h(x')) < \varepsilon_y$. SDPP projects the high dimensional data $\{x_1, x_2, \ldots, x_n\}$ into a lower dimensional space Z with dimensionality $r \ll m$ through the linear function $f : R^m \to R^r$ defined by $f(x) = W^T x, \forall x \in R^m$, where the transformation matrix $W \in R^{m \times r}$. The idea of SDPP is to project the data in such a way that the local geometrical structure of the lower-dimensional subspace preserves the geometrical characteristics of the response space.

Suppose we have n data points $\{x_1, x_2, \ldots, x_n\}, x_i \in R^m$ and their responses $\{y_1, y_2, \ldots, y_n\}$

$$F(W) = \frac{1}{n} \sum_{i=1}^{n} \sum_{x_j \in N(x_i)} (d_{ij}^2(W) - \delta_{ij}^2)^2$$

where, $N(x_i)$ denotes a neighborhood of x_i, $d_{ij}^2(W)$ represents Euclidean metric $d_{ij}^2(W) = \|z_i - z_j\|^2$ that is used to characterized the pairwise distances in Z space and δ_{ij} has the following form:

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \sim j(x_i \text{ and } x_j \text{ belongs to same class}) \\ 1 & \text{otherwise,} \end{cases}$$

Locality around any point x_j is controlled by its k nearest neighbors in $N(x_j)$ where the number k is hyperparameter of SDPP that has to be set beforehand or tuned from data. The value of k is selected by a continuity measure in SDPP[4].

The schematic illustration of SDPP is given in Fig. 2. For a point x in input space, consider three nearest neighbor $N(x) = \{x_1, x_2, x_3\}$. Suppose in output space the neighborhood of y is $\{y_1, y_3, y_4\}$ ie. y_2 is outside of the neighborhood of y. SDPP seeks for the transformation matrix W for which $z_2 = f(x_2)$ is moved outside the neighborhood in the Z-space while z_4 is moved inside to match the local geometry in the Y space as shown in Fig. 2..

Thus, SDPP incorporates a neighborhood graph G_{ij} in the objective function defined as follows:
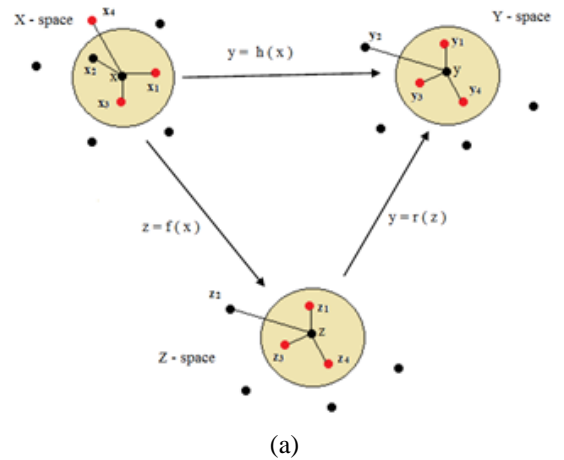
$$G_{ij} = \begin{cases} 1 & \text{if } i \sim j(k - NN neighbor) \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the objective of SDPP is to minimize

$$F(W) = \frac{1}{n} \sum_{i=1}^{n} \sum_{x_j \in N(x_i)} (d_{ij}^2(W) - \delta_{ij}^2)^2$$

which can be written equivalently as follows:

$$F(W) = \frac{1}{n} \sum_{ij} G_{ij}(d_{ij}^2(W) - \delta_{ij}^2)^2$$
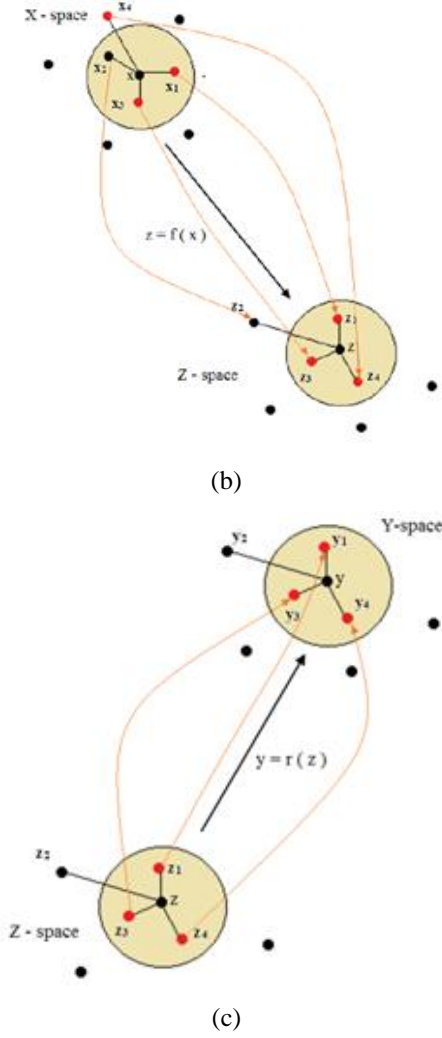

(a)

(b)

(c)

Fig. 2. (a) SDPP: Solid lines indicate connection between neighbors. (b-c) Preservation scheme of the local geometry by SDPP.

IV. Regularized Supervised Distance Preserving Projection (RSDPP)

RSDPP[5] is a modified version of the SDPP method, called Regularized Supervised Distance Preserving Projection (RSDPP proposed by Alencar *et al*.   It relies on the regularization theory. The method has achieved good results in comparison to the state-of-the-art methods in nonlinear dimensionality reduction based on experimental evaluation. In RSDPP, a regularization term is introduced in the SDPP formulation by modifying the cost function. In doing so, L2 norm regularization is chosen and the cost function is rewritten as:

$$J(W) = \frac{1}{n}\sum_{ij} G_{ij}(d_{ij}^2(W) - \delta_{ij}^2)^2 + \lambda vec(W)^T vec(W)$$

where, λ is the regularization parameter and $vec(\cdot)$ is an operator that concatenates the columns of a matrix into a vector. The objective function with $D_{ij} = d_{ij}^2(W)$ and $\Delta_{ij} = \delta_{ij}^2$ can be written as:

$$J(W) = \frac{1}{n}\sum_{ij} G_{ij}(D_{ij} - \Delta_{ij})^2 + \lambda vec(W)^T vec(W).$$

The gradient with respect to W is equal to ∇

$$\nabla_W J = \frac{4}{n}P^T(S - R)PW + 2\lambda W.$$

Each row of P is a data point x_i, S is a diagonal matrix with $S_{ii} = \sum_j R_{ij}$, R is the symmetric matrix with $R = Q + Q^T$ where, $Q = G \odot (D - \Delta)$. To optimize the objective function of both SDPP and RSDPP, conjugate-gradient (CG) optimization[3] technique is considered.

V.   Experiment and Results

We have performed an empirical evaluation of RSDPP for the facial expression recognition model. A Comparison of RSDPP is made in terms of recognition rate with SDPP and some other dimension reduction methods such as SPCA and Eigenface.

Eigenface: Eigenface method[9,10] uses Principal Component Analysis (PCA)[11,12] to reduce the dimension of the image space by maximizing the total scatter of all projected samples. If the total variance matrix S is defined by: $S = \sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^t$ where $\mu \in \mathbb{R}$ is the mean image of all samples and $S \in S^n$, then the basic idea of the Eigenface method is to determine the transformation matrix W in such a way that the determinant of the total scatter matrix $W^T S W$ of the projected sample is maximized.

Supervised principal component analysis (SPCA): Supervised principal component analysis [13](SPCA) is based on the Hilbert-Schmidt independence criterion. It is a generalization of Principal Component Analysis (PCA) that aims to estimate the sequence of principal components that have maximal dependence on the response variable. In SPCA, eigen-decomposition of the weighted covariance matrix enhanced by the kernel of responses is used. Thus, SPCA has closed-form solution like PCA and does not suffer from high computational complexity.

Data description

We have applied the algorithms on two facial data set Yale and JAFFE.

The Yale[14] face database contains 165 grayscale images of 15 individuals with 11 images per subject. We have re-labelled the dataset taking 5 images of each person with 5 different expressions (normal, sad, happy, wink, surprise). So, the resulting data set has 75 images with 15 images of each class.

Sample of some images of Yale faces is shown in Fig. 3.

Neutral   Sad   Happy   Surprise   Wink

Fig. 3. Sample of some images of Yale faces.

In JAFFE[15] data set, there are 213 images of 7 facial expressions (neutral, happy, sad, surprise, anger, disgust, fear) posed by 10 Japanese female models with 3-4 images

per expression. Fig. 4 gives several images of different expressions of JAFFE data.

Angry   Disgust   Fear   Happy   Neutral   Sad   Surprise

Fig. 4. Sample of some images of JAFFE faces.

Numerical Results

The algorithm RSDPP has been implemented on different sizes of training set and testing sets. The recognition rate varies with number of training samples.

Table 1 represents the confusion matrix for Yale dataset when 80% data is selected for training purpose and 20% for testing. For classification, k-nearest neighbour (k-NN)[11,12] is used. It is observed that 1-NN gives the best result. From the experimental results documented in Table1, it can be seen that happy faces have the best recognition rate of 95.2 %, Sad and Neutral faces also have a good recognition rate of 93.4% and 92% respectively. Surprised faces and faces having wink has a fair recognition rate of 88.2% and 89.1% respectively. An interesting fact is that surprise resembles happiness by 10.1%, Sad face is recognized as Neutral by 5.3%. whereas the converse rate is 8%.

Table 1. Confusion Matrix of Yale dataset obtained by RSDPP

		Predicted Expression				
	Expressions	RSDPP	Surprise	Wink	Happy	Sad
Original expression	Neutral	92%	0%	0%	0%	8%
	Surprise	0%	88.2%	1.7%	10.1%	0%
	Wink	4.2%	5.4%	89.1%	1.3 %	0%
	Happy	0%	4.8%	0%	95.2%	0%
	Sad	5.3%	0%	0%	0%	93.4%

Table 2 gives the recognition rate of different expressions by each of the algorithms RSDPP, SDPP, Eigenface, and SPCA. It can be observed that each of the methods except Eigenface has a high recognition rate of happy faces whereas the lowest measure is obtained for faces posed as surprised. Interestingly, most of the misclassified surprised faces were actually predicted as happy faces as in the case of RSDPP shown in the confusion matrix (Table 1). The Eigenface method predicts sad and neutral faces better than other expressions. SPCA gives the lowest prediction rate among these methods for most of the expressions. The results are well represented using bar diagram shown in Fig. 5.

Fig. 6 gives the average recognition rate of all the expressions of Yale data set, which shows that RSDPP outperforms all other methods. The error bar represents the lowest rate and highest rate of prediction for each of the methods. For example, for RSDPP, the lowest value of the error bar is 88.2 which is the prediction rate of surprised faces and the highest value is 95.2 which is the recognition rate of happy faces.

Table 2. Recognition rate of expressions of Yale data by the 4 algorithms.

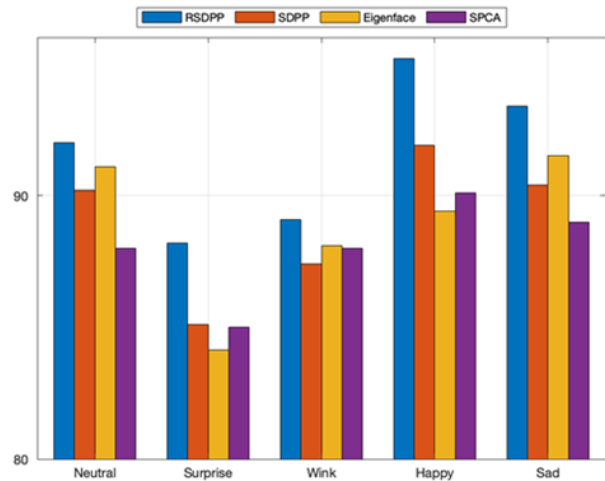| Expressions	RSDPP	SDPP	Eigenface	SPCA
Neutral	92%	90.2%	91.1%	88%
Surprise	88.2%	85.1%	84.15%	85%
Wink	89.1%	87.4%	88.1%	88 %
Happy	95.2%	91.9%	89.4%	90.1%
Sad	93.4%	90.4%	91.5%	89%

Fig. 5. Recognition rate of different expressions of Yale face data by the 4 algorithms.
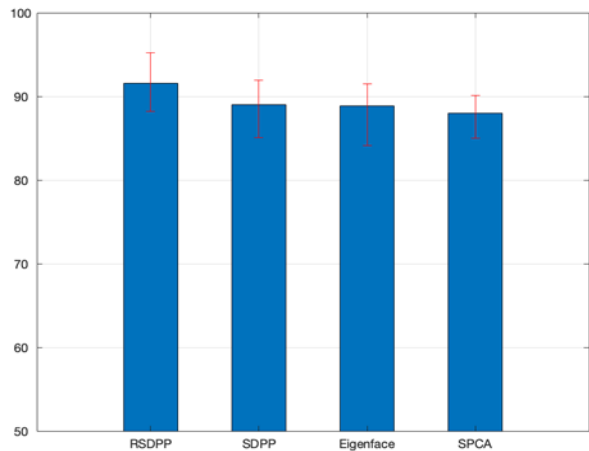

Fig. 6. Average recognition rate of the expressions of Yale data with error bar. Here the lowest value and highest value of the error bar represents the corresponding rate of recognition of expressions.

Similarly, for JAFFE data set, Table 3 represents the confusion matrix[16] for RSDPP. As in the case of Yale data set, happy faces of JAFFE data have the highest recognition rate of 96.4% and the faces that posed for the fear has the lowest rate of prediction. It can be seen that the expressions Anger resembles Disgust and Fear in several cases, 11.5% of the disgust faces and 10.2% of faces with fear pose are predicted as Anger. Similarly, 10.7% of the anger faces are predicted as disgust. Some samples of the faces with original expression and the predicted one are given in Table 5.

Table 3. Confusion Matrix of JAFFE dataset obtained by RSDPP

		Predicted Expression						
	Expressions	Neutral	Happy	Sad	Surprise	Anger	Disgust	Fear
Original expression	Neutral	92%	0%	6%	0%	0%	2%	0%
	Happy	0%	96.1%	0%	3.9%	0%	0%	0%
	Sad	2.1%	0%	94.4%	0 %	0%	0%	3.5%
	Surprise	0%	7.8%	0%	92.2%	0%	0%	0%
	Anger	5.3%	0%	0%	0%	83.4%	10.7%	5.9%
	Disgust	0%	0%	0%	0%	11.5%	85.7%	2.8%
	Fear	0%	0%	8.5 %	0%	10.2%	0%	81.3%

The recognition rates of different expressions of JAFFE data obtained by the four algorithms are documented in Table 4. It can be observed that each of the methods has predicted happy faces better than other expressions whereas the lowest measure is obtained for Fear posed faces.

The results are well represented using bar diagram shown in Fig. 7.

Table 4. Recognition rate of expressions of JAFFE data by the 4 algorithms.

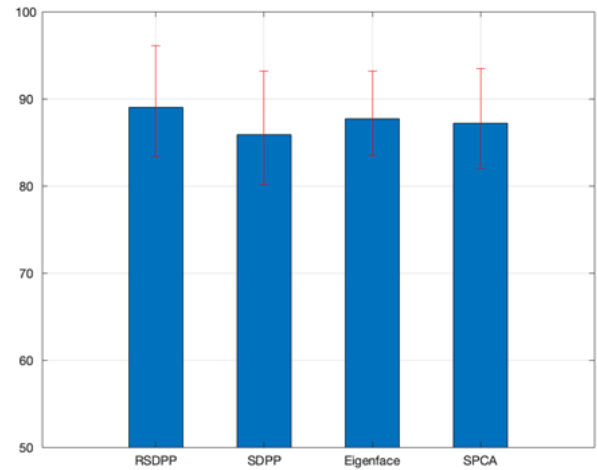| Expressions	RSDPP	SDPP	Eigenface	SPCA
Neutral	90%	87.2%	89.1%	88%
Happy	96.1%	93.2%	93.14%	93.5%
Sad	94.4%	92.4%	91.1%	90 %
Surprise	92.2%	89.2%	91.0%	90.1%
Anger	83.4%	80.2%	83.5%	82%
Disgust	85.7%	81.2%	84.5%	84.7%
Fear	81.3%	77.5%	81.5%	81.2%

Fig. 8. Average recognition rate of the expressions of JAFFE data with error bar. Here the lowest value and highest value of the error bar represents the corresponding rate of recognition of expressions.

Table 5. Sample of images of JAFFE data: Expressions by predicted by RSDPP

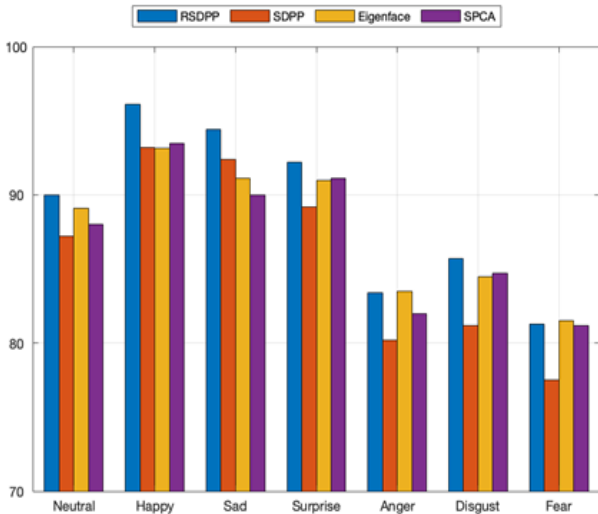| Image	Originally labelled Expression	Predicted expression by RSDPP
	Surprise	Happy
	Happy	Happy
	Anger	Anger
	Disgust	Anger

Fig. 7. Recognition rate of different expressions of JAFFE by the 4 algorithms.

The average recognition rate of the 4 methods with error bar is given in Fig. 8. It shows that RSDPP works better than all other methods with 89% recognition. Eigenface method have 87.69%. For JAFFE data set the average performance of SPCA data is 87.21% which is similar to the Eigenface method

VI. Conclusion

In recent years, researchers are growing their interest in improving expression recognition techniques since it is a key technology of advanced human-computer interaction. In most of the cases the recognition of an image mainly depends on the extracted features. Therefore, feature extraction is a crucial step of the facial expression recognition process. Researchers are using various dimension reduction techniques to extract adequate features. In this paper, we have discussed recently proposed Supervised Distance Preserving Projection (SDPP) and it's a regularized version RSDPP to do the extraction of effective features.

Supervised Distance Preserving Projection method projects high dimensional inputs onto a low-dimensional sub-space. Though, SDPP works well for large data set, Regularized SDPP (RSDPP)aims to deal with ill-conditioning matrix problems such as problems with small samples and high dimensions such as image identification data.

We have implemented the algorithms on two benchmarking face data Yale and JAFFE. The performance of RSDPP is analysed by comparing this method with SDPP and two other state-of-the-art dimensionality reduction methods: SPCA, Eigenface. On the basis of the experimental results, we found that the RSDPP can extract informative features efficiently which leads to a better recognition rate of the expressions.

Reference

1. Huang, Y., F. Chen, S. Lv and X. Wang, 2019. Facial Expression Recognition: A Survey, *Symmetry*, **11**(10), 1189; doi.10.3390/sym11101189.

2. Cowie, R., E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J.G. Taylor, 2001. Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag.*, **18**, 32-80.

3. Mehrabian, A., 1971. Silent Messages (1st ed.). Belmont, CA: Wadsworth. ISBN 0-534-00910-7.

4. Zhu, Z., T. Simila and F. Corona, 2013. Supervised distance preserving projection, *Neural Processing Letters*, **38**, 445-463.

5. Alencar, A. S. C.*et al.*, 2014. Regularized Supervised Distance Preserving Projections for Short-Text Classification,*Brazilian Conference on Intelligent Systems*, Sao Paulo, 216-221, doi: 10.1109/BRACIS.2014.47.

6. Jahan, S., 2018. On Dimension Reduction Using Supervised Distance Preserving Projection for Face Recognition, *Universal Journal of Applied Mathematics*, **6**(3), 94-105,doi: 10.13189/ujam.2018.060303.

7. Jahan, S., 2020. Supervised Distance Preserving Projection using Alternating Direction Method of Multipliers, *Journal of Industrial and Management Optimization, AIMS*, **16**(4),1783-1799. doi: 10.3934/jimo.2019029.

8. Coronaa, F., Z. Zhu, Jr. Souza, M. Mulasd, A. H. d., Muruf, E., Sassuf, L., Barretob, G. and Baratti, R., 2015. Supervised Distance Preserving Projections: Applications in the quantitative analysis of diesel fuels and light cycle oils from NIR spectra, *Journal of Process Control*,**30**, 10–21.

9. Turk, M. and Pentland, A., 1991. Face recognition using eigenfaces, *In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, 586–591.

10. De, A., S. Saha, M. C. Pal, 2015. A Human Facial Expression Recognition Model based on Eigen Face Approach, *International Conference on Advanced Computing Technologies and Applications (ICACTA-2015), Procedia Computer Science*,**45**, 282-289.

11. Theodoridis, S. and K. Koutroumbas, 2009. Pattern Recognition. Elsevier Inc.

12. Theodoridis, S. and K. Koutroumbas, 2010. An Introduction to Pattern Recognition, A MATLAB approach, *Elsevier Inc*.

13. Barshan, E., A. Ghodsi, Z. Azimifar and M. Z. Jahromi, 2010. Supervised principal componentanalysis: visualization, classification and regression on subspaces and submanifolds, *Pattern Recognition*.**44**, 1357-1371.

14. http://www.cad.zju.edu.cn/home/dengcai/Data /FaceData.html.

15. Michael, J. Lyons et al., 1998. The Japanese Female Facial Expression (JAFFE) Dataset [Data Set]. Zenodo. http://doi.org/10.5281/zenodo.3451524.

16. Mayya, V. R. M. Pai, M. Pai M. M., 2016. Automatic Facial Expression Recognition Using DCNN, *Procedia Computer Science,***93**, 453-461.