Variants of Locality Preserving Projection for Modular Face and Facial Expression Recognition

Gitam Shikkenawis(✉) and Suman K. Mitra

Dhirubhai Ambani Institute of Information and Communication Technology, Gandhianagar, India
{201221004,suman_mitra}@daiict.ac.in

Abstract. Locality Preserving Projection (LPP) is one of the widely used approaches for finding intrinsic dimensionality of high dimensional data by preserving the local structure. Data points which are neighbors but belong to different classes are thereby projected as neighbors in the projection space, causing problem of discrimination. Various extensions of LPP have been proposed to enhance the discrimination power achieve better between class separation. In case of face recognition using full face images, if any portion of the face image is distorted, it may reflect on the recognition performance. Humans have the capability to recognize faces even by looking at some parts of the face. This article is an attempt to replicate the same on machines by only considering some of the informative regions of the face. Instead of the entire image, variants of LPP are applied on parts of face images and recognition is performed by combining the results of their reduced dimensional representations. Face and facial expression recognition experiments have been performed on some of the benchmark face databases.

1 Introduction

Recognizing human faces and expressions comes naturally to humans even under adverse viewing conditions such as various lighting conditions, viewing angles, poses, expression and appearance changes, occlusions etc. Though, significant advances have been achieved in last few decades in the area of face recognition especially in constrained environment, a face recognition system as good as the Human Visual System (HSV) is yet to be achieved. Feature based techniques mark prominent features from faces such as eyes, nose, mouth and compare the test images based on selected group of features [3,5,6]. On the other hand, appearance based techniques work on the idea that high dimensional face images often belong to intrinsically lower dimensional manifold and can be represented using very few coefficients. Such approaches, generally known as dimensionality reduction (DR) approaches, Principle Component Analysis (PCA) [13], Linear Discriminant Analysis (LDA) [1], Locality Preserving Projections (LPP) [4] etc. transform high dimensional face data into significantly lower dimensions and perform recognition task have become very popular.

© Springer International Publishing AG 2017
B.U. Shankar et al. (Eds.): PReMI 2017, LNCS 10597, pp. 141–147, 2017.
https://doi.org/10.1007/978-3-319-69900-4_18

The property of preserving local information make LPP one of the most popular DR techniques to be used for face recognition lately. Various extensions of LPP to make it more robust and suitable to face recognition have been proposed [2,10,11,14]. So far, these DR approaches have been applied on full face images. In such cases, if any portion of face image is distorted, it may reflect on the recognition performance. Also, it has been observed that even after looking at some of the informative regions of the face such as eyes, nose and lips, humans can easily recognize the person. Hence, a more robust face recognition system can be developed by combining the feature based and appearance based techniques. Instead of whole face images, some of the specific informative regions of a face can be extracted and DR techniques can be applied only on the extracted parts from the face.

In one such approach, PCA is performed on the nose and eyes of the face images [8]. A modular PCA based approach [9] divides the face image in smaller parts and then PCA is applied on these portions separately. As face regions are considered for recognition, variations in expressions or pose or illumination in the image will affect only some part of the image, hence a better recognition rate can be expected. Modular Locality Preserving Projection (MLPP) is proposed in this article, which takes the local regions such as eye, nose and lips of a face as input of the DR approach separately and produces final result by fusing the outcome of these regions. In particular, utilization of Extended Locality Preserving projection (ELPP) [10] and Locality Preserving Discriminant Projection (LPDP) [11] for the proposed modular face and expression recognition is explored in this article. Suitability of the proposal is tested on databases having expression variation.

Organization of the paper is as follows: Sect. 2 discusses variants of LPP. Modular Locality Preserving Projection (MLPP) that works on some prominent regions of face images is explained in Sect. 3 along with face and expression recognition experiments.

2 Variants of Locality Preserving Projection

As discussed in Sect. 1, in this article, variants of LPP i.e. Extended Locality Preserving Projection (ELPP) [10] and Locality Preserving Discriminant Projection (LPDP) [11] have been used for modular face recognition.

Extended Locality Preserving Projection
Extended LPP [10] is an extension of LPP [4] towards making it more robust and enhance the DR capability. LPP [4] emphasizes on the local structure of the data to preserve the neighborhood information. Due to the property of LPP to depend only on a few nearest neighbors, ambiguity may arise as a result of adjacency of data points from different classes. ELPP not only extends the neighborhood to a moderate distance from the point of interest, it also tries to explore natural grouping of the data with the use of k-means clustering.

The goal of ELPP is, data points that are neighbors in the high dimensional space should continue to remain neighbors in the lower dimensional space as

well. Transformation matrix $\mathbf{w}$ to represent data in the lower dimensional space is obtained by solving the generalized eigenvalue problem: $\mathbf{XLX^T w} = \lambda \mathbf{XMX^T w}$; here, $\mathbf{X}$ is the data matrix, $\mathbf{L}$ is the Laplacian matrix i.e. $\mathbf{L} = \mathbf{M} - \mathbf{S}$. $\mathbf{S}$ is similarity matrix that takes care of neighborhood information and $M_{ii} = \sum_i S_{ij}$. Data points clustered in one class using k means clustering are considered neighbors and assigned weight in $\mathbf{S}$. Weighing is performed using a monotonically decreasing function that weighs the neighboring data points depending on the distance between them [10]. This choice of weight makes sure that neighboring data points remain neighbors in the newly obtained ELPP subspace as well.

Locality Preserving Discriminant Projection
Though ELPP tries to resolve the ambiguity arising due to closeness of data points belonging to different classes, no emphasis is given by LPP and ELPP to enhance the between class discriminating power. LPDP [11], in addition to inheriting the properties of ELPP of preserving the similarity information, tries to discriminate data points from different classes by taking into consideration the dissimilarity information as well. The aim is to achieve better class separation by using weighing functions for both similarity and dissimilarity of the data points. The generalized eigen value problem thus turns out to be: $(\mathbf{XL_S X^T} - \mathbf{XL_D X^T})\mathbf{w} = \lambda \mathbf{XX^T w}$; as in case of ELPP, here, $\mathbf{X}$ is the data matrix, $\mathbf{L_S}$ is the Laplacian matrix obtained from the similarity matrix i.e. $\mathbf{L_S} = \mathbf{M_S} - \mathbf{S}$. Similarity matrix $\mathbf{S}$ is computed in the same manner as that of ELPP and $MS_{ii} = \sum_i S_{ij}$. Data points that belong to different classes are considered dissimilar and weights in $\mathbf{D}$ are assigned to ensure maximum class discrimination [11] in a monotonically increasing fashion. $\mathbf{L_D} = \mathbf{M_D} - \mathbf{D}$ and $MD_{ii} = \sum_i D_{ij}$. Thus, in addition to preserving the local information, LPDP also takes into account the dissimilarity between data points to achieve enhanced class discrimination.

In this article, ELPP and LDPD have been used to reduce the dimensionality of data. The main contribution of this article is use of DR technique only on some of the informative regions of the face image instead of full face as discussed in next section.

3 Modular Locality Preserving Projection

Dimensionality reduction methods, when applied for full face images, may not work as expected, if any portion of the face image is distorted. Any obstacles, changes in the facial expressions or pose may also degrade the performance. Lower dimensional representations of the regions which are not affected by changes will match with that of the same individual's face regions in normal conditions. Hence, it is expected that the recognition results can be improved by applying the DR techniques on local face regions separately. It seems that eyes, nose and lips are more informative for identifying a person. By having a look only at one of these face parts the person can be identified. Suitability of ELPP and LPDP for identifying faces and facial expressions is tested by applying it locally on the faces. In particular, dimensionality reduction is applied on

significant regions such as eyes, nose and lips. Here, these parts have been cut manually from the face regions. The process of extracting the regions from face can be also automated by first detecting the eyes [5] and then using the golden ratio to cut other informative regions.

Intrinsic dissimilarity between pair of eyes of two different persons is hard to be identified by machine as there will be a lot of overlap between eye regions of different persons. ELPP is expected to find out this dissimilarity in a little better way than LPP as it performs much better in the overlapping regions. However, as LPDP takes into consideration both similarity and dissimilarity information while obtaining the basis, it should be able enhance the discrimination ability and result in improved recognition performances. As we are moving towards more local areas of the face and then applying LPP, ELPP and LPDP on these local regions, this method is called Modular Locality Preserving Projection (MLPP). Two different modular approaches are analyzed here.

Modular Approach #1
In the first approach, eyes, nose and lip regions are considered separately; LPP, ELPP and LPDP are applied on these regions and classification using different parts is carried out. Clustering experiments are performed on Video database [12] containing face images of 11 persons having four different expressions and the Japanese Female Facial Expression JAFFE database [7] are used. All the face images are cut as shown in Fig. 1. Only the eyes, nose and lip regions are extracted from the whole image. The results of clustering the projected data using LPP, ELPP and LPDP using different number of dimensions are shown in Table 1.

Fig. 1. Examples of the selected face regions used from Video database for modular approach #1 (Left) modular approach #2 (right)

From the results, it can be concluded that eyes are the most informative and discriminative portions of the human face. All three approaches are able to discriminate between different person's eyes. On the Video database, with LPP, almost all the dimensions are required for discrimination, whereas the other two approaches are doing it using much less dimensions, thus enhancing the reducibility capacity. For JAFEE database also, more than 96% accuracy is achieved using the all three approaches. For only nose portion, the accuracies are higher than 80% for both the databases. In Video database, results on lip regions are not that encouraging because of a lot of lip variation in the database with LPP and ELPP but LPDP is performing very well, achieving almost 100%

Table 1. Results (%) of clustering eyes, nose and lip regions separately from the Video and JAFFE database using nearest neighbor approach.

DA-IICT	# Dimensions														
	2			10			50			500			MAX		
	LPP	ELPP	LPDP	LPP	ELPP	LPDP	LPP	ELPP	LPDP	LPP	ELPP	LPDP	LPP	ELPP	LPDP
Eyes	11.3	74.8	99.52	20.5	99.45	**100**	18	100	100	21.46	100	100	97.82	100	100
Nose	16.91	84.55	99.16	18.55	91.64	**100**	14.91	91.64	100	15	91.82	100	88.55	92	100
Lips	20.19	18.55	98.40	20	17.10	**100**	19.82	19.20	100	19.20	21.82	100	19.82	22.55	100
JAFFE	# Dimensions														
	2			10			50			150			MAX		
	LPP	ELPP	LPDP	LPP	ELPP	LPDP	LPP	ELPP	LPDP	LPP	ELPP	LPDP	LPP	ELPP	LPDP
Eyes	13.16	41.53	73.68	23.16	89.48	91.58	25.27	94.22	93.68	59.43	95.79	**96.84**	96.32	96.32	96.84
Nose	6.85	43.11	63.68	13.79	81.58	75.26	13.16	87.37	83.16	47.37	85.79	85.26	86.32	**87.8**	85.26
Lips	14.22	27.9	60.50	14.22	69.43	84.00	10	82.11	85.00	47.9	84.74	85.00	82.11	85.27	**85.50**

accuracy. In case of JAFFE database, where there is a lot of expression variation, much better results are obtained for all the three DR approaches using only lip region.

The experimental results show that in this manner, the faces could even be recognized easily from eyes only. However, joining the decisions of face recognition separately from these regions is yet to be explored. In cases such as video database where there is a lot of lip variation because of expressions, if full face images are considered, variation in expressions can cause problems for recognition tasks. On the other hand, as suggested here, if the regions are considered separately, faces could easily be recognized using only the eyes and nose regions. Hence, by combining the decisions of these face regions; a robust face recognition system can be designed.

Modular Approach #2

During the task of expression analysis, we observed that apart from eyes, nose and lips/mouth, forehead also plays very important role as far as expressions are concerned. Hence, in the second approach, these four portions from the face image are used for recognition purpose. The regions cut from the face image are shown in Fig. 1.

Unlike the first approach discussed earlier, here, all portions are combined together in vector format to form a data point. The data points generated this way undergo dimensionality reduction and are classified using nearest neighbor approach. This approach is tested on the video database for both face and expression recognition. It is also to be noted that for expression recognition, expression labels of the data points are considered to be known, hence the neighbors are decided based on the class labels. These can be considered as the supervised variants of LPP and ELPP. Face recognition results with varying dimensions using LPP, ELPP and LPDP are reported in Table 2. It can be observed that LPP, ELPP and LPDP perform extremely well achieving almost 100% recognition accuracy using only 10 dimensions. With 2 most significant dimensions, LPDP surpasses both LPP and ELPP.

The Video database mainly contains four facial expressions for each subject namely normal, happy (laughing), angry and shock. Expression recognition

Table 2. Face recognition accuracy (in %) on Video database using LPP, ELPP and LPDP

# Dimensions	2	10	20	30	40	50
LPP	87.65	**100**	100	100	100	100
ELPP	93.85	99.85	100	100	100	100
LPDP	97.15	**100**	100	100	100	100

experiments have been performed in two different ways: (1) Randomly selecting training and testing samples from the same set of persons, (2) Randomly selecting training and testing samples from different persons i.e. testing set contains face images of the persons that have not been included for training. This liberty can be taken as we are recognizing the expressions and training set contains similar expressions for other persons. This exercise makes expression recognition more challenging.

When the training and testing samples are randomly selected from the same set of persons for LPP, ELPP and LPDP, more than 98% expression recognition accuracy is attained using ELPP and LPDP with only 40 strongest dimensions as opposed to 12000 dimensions of the raw data points in the original space as shown in Table 3.

Table 3. Expression recognition accuracy (in %) on Video database using LPP, ELPP and LPDP

# Dimensions	Same persons for training, testing						Different persons for training, testing					
	2	10	20	30	40	50	2	10	20	30	40	50
LPP	79.25	94.9	96.3	97.25	97.65	97.8	66	76.50	82.25	80.50	77.75	79.00
ELPP	79.40	97.45	98.05	98.60	**98.65**	98.65	71.25	81.00	83.00	83.25	85.25	88.00
LPDP	93.10	95.85	96.75	97.45	98.2	98.45	85.00	93.75	94.00	94.00	94.00	**94.25**

In practical scenarios, it is not possible to have training data for all the test samples whose expressions are being recognized. A similar experiment, where the person whose expressions are to be recognized has not been included in the training set, is performed. Though the recognition rate has reduced, 88% accuracy has been achieved using ELPP with only 50 dimensions. On the other hand, LPDP is able to produce 94% recognition rate with only 20 most significant dimensions. Though ELPP performs better than LPP with less number of dimensions, in most of the scenarios, LPDP surpasses both LPP and ELPP in terms of both recognition accuracy and reducibility capacity.

The initial set of experiments reported in this article suggest that the modular approaches using only some prominent portions of face images can be useful for face and expression recognition. The idea needs to be further explored for other databases having distortions and occlusions in the face image.

4 Conclusion

Capability of ELPP and LPDP to recognize a person using partial information from the whole face image is explored in this work. Dimensionality reduction is applied on most informative regions of the face i.e. eyes, nose and lips. It is observed that only eyes are significant enough to distinguish faces of different persons in most of the cases, however, by fusing the results of different face parts, a more robust face recognition system can be deployed. In addition to face recognition, expression recognition results also support the argument of using only informative regions from face images for recognition tasks. Thus, the modular approach suggested in this article can further be applied for face and expression recognition task to attain more robust results specially for challenging databases having distorted or occluded face images.

References

1. Belhumeur, P., Hespanha, J., Kriengman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Patt. Anal. Mach. Intell. **19**(13), 711–720 (1997)
2. Dornaika, F., Assoum, A.: Enhanced and parameterless locality preserving projections for face recognition. Neurocomputing **99**, 448–457 (2013)
3. Govindaraju, V., Srihari, S.N., Sher, D.B.: A computational model for face location. In: International Conference on Computer Vision, pp. 718–721 (1990)
4. He, X., Niyogi, P.: Locality preserving projections. In: NIPS (2003)
5. Lam, K., Yan, H.: Locating and extracting the eye in human face images. Patt. Recogn. **29**(5), 771–779 (1996)
6. Li, S., Jain, A.: Hand Book of Face Recognition. Springer, New York (2011)
7. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: Automatic Face and Gesture Recognition, pp. 200–205. IEEE Computer Society (1998)
8. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 84–91 (1994)
9. Gottumukkal, V.K.A.R.: An improved face recognition technique based on modular PCA approach. Patt. Recogn. **25**, 429–436 (2004)
10. Shikkenawis, G., Mitra, S.K.: Improving locality preserving projection for dimensionality reduction. In: Emerging Applications of Information Technology, pp. 161–164. IEEE Computer Society (2012)
11. Shikkenawis, G., Mitra, S.K.: Locality preserving discriminant projections. In: IEEE International Conference on Identity, Security and Behavior Analysis (ISBA) (2015)
12. Shikkenawis, G., Mitra, S.K.: On some variants of locality preserving projection. Neurocomputing **173**(2), 196–211 (2016)
13. Turk, M., Pentland, A.: Eigenfaces for face recognition. Cogn. Neurosci. **3**(1), 71–86 (1991)
14. Xu, Y., Zhong, A., Yang, J., Zhang, D.: LPP solution schemes for use with face recognition. Patt. Recogn. **43**(12), 4165–4176 (2010)