

2D-LPI: Two-Dimensional Locality Preserving Indexing

S. Manjunath¹, D.S. Guru¹, M.G. Suraj¹, and R. Dinesh²

¹ Department of Studies in Computer Science, University of Mysore, Mysore, India

² Honeywell Technology Solutions, Bengaluru, India

Abstract. In this paper, we present a new model called two-dimensional locality preserving indexing (2D-LPI) for image recognition. The proposed model gives a new dimension to the conventional locality preserving indexing (LPI). Unlike the conventional method the proposed method can be applied directly on images in 2D plane. In order to corroborate the efficacy of the proposed method extensive experimentation has been carried out on various domains such as video summarization, face recognition and fingerspelling recognition. In video summarization we compare the proposed method only with 2D-LPP which was recently used for video summarization. In face recognition and fingerspelling recognition we compare the proposed method with the conventional LPI and also with the existing two-dimensional subspace methods viz., 2D-PCA, 2D-FLD and 2D-LPP.

1 Introduction

Latent semantic indexing (LSI) was proposed in [1] to cluster text documents. LSI finds the best subspace which approximates the original document space onto a smaller subspace by minimizing the global reconstruction error. Initially LSI was used for text clustering / classification and later it was explored to index audio documents [2], image retrieval [3] and video content modeling [4]. In 1999, Kurimo [2] developed a method to enhance the indexing of spoken documents with the help of LSI and self organizing map. Zhu and Grosky [3] extracted global color histogram from images and latent semantic analysis was performed to learn those extracted features to develop an image retrieval system. In [4], Souvannavong et al., used latent semantic indexing for modeling video content with Gaussian mixture model. Although LSI finds an appropriate subspace to obtain lower dimensional representation of original space, it is found that LSI seeks to uncover the most representative features rather than the most discriminative features for representation [5].

Hence, Locality Preserving Indexing (LPI) was developed for document representation [5] on the basis of Locality Preserving Projection [6] which preserves local structure of the given high dimensional space in lower dimension using spectral clustering. Later, it was modified by introducing an additional step of singular value decomposition in [7] to original LPI method [5] in order to ensure that the transformed matrix is of full order. The modified LPI [7] tries to ascertain both

geometric and discriminating structure of the document space by discovering the local geometrical structures [7]. In order to project images onto the LPI subspace, images need to be transformed into one dimensional vector thereby requiring more memory and high computational time. However, a number of works are reported in the literature where vector based projections techniques are successfully extended to image based projection techniques [8][9][10]. The image based projection techniques project images directly onto lower dimensional space rather than transforming them into one dimensional vector [10], thereby working directly on matrices (images) and hence requiring less memory and less computational time. Motivated by the works of [8][9][10], in this paper we extend the conventional LPI [7] to two dimensional LPI, which can be applied directly on images to project them onto lower dimensions by preserving both representative as well as discriminative features of images. From the experimentation it is found that 2D-LPI is more effective as well as efficient when compared to the conventional LPI and also further study resulted that proposed 2D-LPI is better than the existing 2D-LPP [10], 2D-PCA [8] and 2D-FLD [9].

2 Two-Dimensional Locality Preserving Indexing (2D-LPI): Proposed Model

Consider a set of N images $I_1, I_2, I_3, \dots, I_N$ each of $(m \times n)$ size. We propose a method that maps the original $(m \times n)$ dimensional space onto $(m \times d)$ dimension with $d \ll n$ by projecting each image I_i onto a subspace using W , a $(n \times d)$ dimensional transformation matrix as $Y = IW$ where Y is $(m \times d)$ dimensional projected features of image I .

2.1 The Algorithm

The proposed algorithm is as follows

1. **Construct the graph adjacency matrix:** Let G denote a graph with images as nodes. Image I_i is connected to image I_j if and only if I_j is one among the k -nearest neighbors of image I_i .
2. **Choose the weight:** Let S be the similarity weight matrix given by

$$S_{ij} = \begin{cases} \frac{\exp\{-\|I_i - I_j\|\}}{t} & \text{If image } I_i \text{ is connected to } I_j \text{ of } G \\ 0 & \text{Otherwise} \end{cases} \quad (1)$$

where t is a suitable constant.

The weight matrix S of the graph G gives the local structure of the image space. A diagonal matrix D whose entries are column (or row) sum of weight matrix S is computed as $D_{ii} = \sum_i S_{ji}$ and the Laplacian matrix L is computed as $L = D - S$.

3. **Singular Value Decomposition:** Compute the weighted mean matrix of all the images, $\bar{I} = \frac{1}{N}(\sum_{i=1}^N I_i D_{ii})$. Decompose the obtained weighted mean matrix using singular value decomposition (SVD) to obtain matrices U and

V which contain left singular and right singular vectors as their columns respectively along with the diagonal matrix σ .

$$\hat{I} = U\sigma V^T \text{ where } \hat{I} = [\hat{i}_1, \hat{i}_2, \dots, \hat{i}_N]$$

The right singular matrix V is considered as transformation matrix of SVD and it is denoted by W_{SVD} i.e., $W_{SVD} = V$.

4. **Two-dimensional LPI (2D-LPI) Projection:** Compute the eigen vectors and eigen values by solving the generalized eigen problem as shown in (2), where W_{2DLPI} and λ are respectively the eigen vectors and the eigen value.

$$A^T (L \otimes I_m) A W_{2DLPI} = \lambda A^T (D \otimes I_m) A W_{2DLPI} \quad (2)$$

Here, A is an $(mN \times n)$ matrix generated by arranging all the images I in a column, $A = [A_1^T, A_2^T, A_3^T, \dots, A_N^T]^T$. Operator \otimes denotes Kronecker product and I_m represents an identity matrix of order m .

Let $W_{2DLPI} = [w_1, w_2, w_3, \dots, w_d]$ and w_i $i = 1, 2, 3, \dots, d$ be the first d unitary orthogonal solution vectors of (2) corresponding to the d smallest eigen values in the order of $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_d$. These eigen values are nonnegative because both $A^T (L \otimes I_m) A$ and $A^T (D \otimes I_m) A W$ are symmetric $(n \times n)$ matrices and positive semi-definite. Thus W_{2DLPI} constitutes the $(n \times d)$ projection matrix, and the projection is done as follows.

$$Y_i = I_i W, \quad i = 1, 2, 3, \dots, n \quad (3)$$

Where $W = W_{SVD} W_{2DLPI}$ and $Y_i = [Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}, \dots, Y_i^{(d)}]$ is an $(m \times d)$ feature matrix of I_i .

2.2 Image Recognition

In the previous section 2.1, a set of images is projected on to 2D-LPI subspace. The distance between two projected matrices say $Y_i = [Y_i^{(1)}, Y_i^{(2)}, \dots, Y_i^{(d)}]$, $Y_j = [Y_j^{(1)}, Y_j^{(2)}, \dots, Y_j^{(d)}]$ is calculated using Euclidean distance between the corresponding vectors i.e.,

$$dist(Y_i, Y_j) = \sum_{k=1}^d \left\| Y_i^{(k)} - Y_j^{(k)} \right\| \quad (4)$$

Suppose there are t_n number of training samples T_1, T_2, \dots, T_{t_n} which are projected onto a 2D-LPI space and each of these samples are assigned to a class C . The class C of a test image T^l is identified by the use of nearest neighbor rule based Euclidean distances.

3 Experimentation 1

To study the efficacy of the proposed 2D-LPI we have carried out experimentation on video summarization, face recognition and fingerspelling recognition.



Fig. 1. Manually generated story board

3.1 Video Summarization

Video summarization is a process of representing a complete video with a small set of visually important frames. Video summarization is necessary for easy and fast content accessing and browsing of a large video database. Also it facilitates content based-video indexing and retrieval [11]. To study the efficacy of the proposed 2D-LPI, we conducted an experimentation to summarize a lecture video. While conducting the experimentation we decoded the video into frames. Then each frame is converted from RGB color to gray-level image and resized into 40×60 pixel image size. A lecture video contains 9 shots with 2743 frames at 30 frames per second. A story board of the lecture video is manually created by selecting one or more key frames from each shot as shown in Fig. 1.

Inorder to create automatic story board all frames of the video is used and projected onto the proposed 2D-LPI model. The number of eigen vectors ($d = 4$) were selected experimentally which gave best results and $K = 5$ is used to construct the adjacency matrix G . The obtained image features are used to cluster the frames using C-means clustering. C-means algorithm was applied 10 times with different starting points, and the best result in terms of the objective function of C-means was recorded. The centroid of each cluster is used as cluster representative which are used in the automatic generated story board and the generated story board using the proposed method is as shown in Fig. 2 (a). Inorder to know effectiveness of the proposed method on video summarization we compare the summarization results obtained by 2D-LPP [11] method, which is used for video summarization recently. Figure 2(b) shows the story board generated using 2D-LPP method.

Table 1 shows the details of the shots used for experimentation, row labeled expert depicts to number of keyframes identified by an expert and row corresponding to the proposed and 2D-LPP depicts the number of keyframes gen-



Fig. 2. (a) Automatic generated story board using proposed 2D-LPI method, (b) Automatic generated story board using proposed 2D-LPP method

Table 1. Shot details and the number of keyframes selected manually and proposed 2D-LPI method and existing 2D-LPP method

Shot	1	2	3	4	5	6	7	8	9
Shot Details	[1-16]	[17-130]	[131-752]	[753-912]	[913-1007]	[1008-1657]	[1658-1847]	[1848-1999]	[2000-2743]
Expert	1	1	2	1	1	4	1	1	3
Proposed	1	1	1	1	1	4	1	1	3
2D-LPP	0	1	1	1	1	1	1	1	4

erated by the proposed method and 2D-LPP method [11]. From the Table 1, it can be seen that the existing 2D-LPP method omits a few keyframes where as the proposed method extracts keyframes approximately equal to the expert. Another important observation to be made is, shot 1 contains only 15 frames which has frames containing gradual transition from shot 1 to shot 2 and hence we have zero number of keyframes of shot 1 by 2D-LPP method. The proposed method extract key frames from shot 1, where as the 2D-LPP merges the shot 1 with shot 2. This type of video was intensionally taken to study the performance of both the methods.

3.2 Experimentation 2

In order to corroborate the efficacy of the proposed method we also conducted experimentation on ORL face, Yale face and fingerspelling datasets [12] useful for sign language recognition. The face images were resized into 40×50 pixels and the first six images of the ORL and Yale databases are used for training and the remaining four images are used for testing with $K = 5$ to construct the adjacency matrix G . The ORL and Yale achieved a maximum recognition rate of 96.88 and 88 percentage using conventional LPI with 25 and 50 dimensions and the proposed 2D-LPI method achieved a maximum recognition rate of 98.75 and 96 percentage for 40×7 dimensions respectively. On fingerspelling dataset conventional LPI achieved 86.88 percentage for 55 dimensions and proposed 2D-LPI achieved a maximum recognition rate of 89.88 percentage for 50×7 dimensions. Also we compare the recognition rate of the proposed method with 2D-PCA, 2D-FLD and 2D-LPP and the results are summarized in Table 2. From Table 2 it can be seen that the proposed method outperforms the other methods on both face and fingerspelling datasets.

Table 2. Maximum recognition rates on face datasets (ORL and Yale) and fingerspelling dataset

	2DPCA	2DFLD	2DLPP	2DLPI
ORL	96.88	97.50	96.25	98.75
Yale	90.67	89.00	88.00	96.00
Fingerspelling	88.75	89.27	72.50	89.38

4 Conclusion

In this paper we have proposed a two dimensional locality preserving indexing (2D-LPI) method suitable for images. The proposed method is more intuitive in handling images and performs better than the conventional LPI and is even better than the 2D-LPP. To study the efficacy of the proposed method, we have carried out experimentation on video summarization, face recognition and fingerspelling recognition. From the experimentation it is evident that 2D-LPI performs better than conventional LPI and also perform better than 2D-LPP, 2D-PCA and 2D-FLD method.

References

1. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. *Journal of the American Society of Information Science* 41, 391–407 (1990)
2. Kurimo, M.: Indexing audio documents by using latent semantic analysis and som. In: *Kohonen maps*, pp. 363–374. Elsevier, Amsterdam (1999)
3. Zhao, R., Grosky, W.I.: From features to semantics: Some preliminary results. In: *Proceedings of International Conference on Multimedia and Expo* (2000)
4. Souvannavong, F., Merialdo, B., Huet, B.: Video content modeling with latent semantic analysis. In: *Proceedings of Third International Workshop on Content-based Multimedia Indexing* (2003)
5. He, X., Cai, D., Liu, H., Ma, W.Y.: Locality preserving indexing for document representation. In: *Proceedings of International Conference on Research and Development in Information Retrieval*, pp. 96–103 (2004)
6. He, X., Niyogi, P.: Locality preserving projections. In: *Advances in Neural Information Processing Systems* (2003)
7. Cai, D., He, X., Han, J.: Document clustering using locality preserving indexing. *IEEE Transactions on Knowledge and Data Engineering* 17, 1624–1637 (2005)
8. Yang, J., Zhang, D., Frangi, A.F., Yu Yang, J.: Two-dimensional pca a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(1), 131–137 (2004)
9. Li, M., Yuan, B.: 2d-lda a statistical linear discriminant analysis for image matrix. *Pattern Recognition* 26(5), 527–532 (2005)
10. Chen, S., Zhao, H., Kong, M., Luo, B.: 2d-lpp a two dimensional extension of locality preserving projections. *Neurocomputing* 70, 912–921 (2007)
11. Xu, L.Q., Luo, B.: Appearance-based video clustering in 2d locality preserving projection subspace. In: *Proceedings of ACM International Conference on Image and Vide Retrieval* (2007)
12. Guru, D.S., Suraj, M.G.: Fusion of pca and fld at feature extraction level for finger spelling recognition. In: *Proceedings of the Third Indian International Conference on Atrificial Intelligence (IICAI)*, pp. 2113–2123 (2007)