Face Recognition using a Color Subspace LDA approach

M. Thomas, C. Kambhamettu
Video/Image Modeling and Synthesis Lab
University of Delaware
Newark, DE
{manivt, chandra}@cis.udel.edu

S. Kumar
Bell Laboratories, Lucent Technologies,
600 Mountain Avenue,
Murray Hill, NJ 07974
senthil@bell-labs.com

Abstract

This paper delves into the problem of face recognition using color as an important cue in improving the accuracy of recognition. To perform recognition of color images, we use the characteristics of a 3D color tensor to generate a color LDA subspace, which in turn can be used to recognize a new probe image. To test the accuracy of our methodology, we computed the recognition rate across two color face databases. We observe that the use of the LDA color subspace significantly improves recognition accuracy over the standard gray scale approach without sacrificing computational efficiency.

1. Introduction

Research in face recognition is currently one of the most important computer vision application ranging from forensics to surveillance. When observing the developmental trend of appearance based face recognition algorithms, we see that there are two possible pathways: the global appearance and the local appearance based algorithms [17]. In the former, the face is considered in a holistic sense and recognition is performed on an entire face image. In the latter case, the face is considered as a combination of local pixel neighborhoods (eyes, nose and mouth) and recognition is performed by combining the local recognition rates over these neighborhoods. Research has shown that the local appearance based recognition performs better than the holistic approach especially under constraints of partial occlusion [6][9]. Global techniques have the single advantage that recognition accuracy is sacrificed for better computational efficiency and ease of implementation, since local models require accurate identification of semantically meaningful pixel neighborhoods.

The development of a face recognition system thus requires a balance between accuracy of estimation and efficiency of computation. In order to satisfy such a need, we propose a global face recognition algorithm, where we use color to improve accuracy. The organization of the paper is as follows. We first describe some of the important works that defined our study, subsequently followed by a description of the algorithm implemented in this study and its performance under the experimental setup. We conclude our work with results attained in our recognition framework and implications of the findings.

2. Background Studies

One of the earliest contributions to the field of recognition via subspace decomposition was by Turk and Pentland [12]. The Principal Component Analysis (PCA) of the face space or the "Eigenfaces" approach was seminal. It has since then given rise to many variants and reformulations that attempt to tackle the difficulties present in its original formulation. A different subspace decomposition strategy was proposed by Belhumeur et al. [1] based on the Fisher's Linear Discriminant Analysis (LDA). This was shown to be superior to the original PCA formulation but required an increased computational requirement. The "Eigenfaces" and the "Fisherfaces" can be considered the two main subspace decomposition work that have been improved upon by other researchers. Comparative studies by Ruiz-del-Solar and Navarrete [3], Delac et al. [4] and G. Shakhnarovich and Moghaddam [11] to name a few, might direct interested readers to the available subspace techniques.

The underlying ideology among the subspace techniques is the maximization of the Rayleigh coefficient [2] subject to scalar constraints on ω.

$$\arg\max_{\omega} \mathcal{L}(\omega) = \arg\max_{\omega} \left(\frac{\omega^T \mathbf{M} \omega}{\omega^T \mathbf{N} \omega} \right) \quad (1)$$

where $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n \times n}$ are two symmetric matrices.

In the case of "Eigenfaces" or one dimensional PCA [12], equation 1 simplifies to the ordinary symmetric eigenvalue problem with $\mathbf{M}$ depicting the total variation of all the data points. In the case of "Fisherfaces" or the one dimensional LDA [1], the subspace decomposition attempts to simultaneously maximize the inter-class variation while minimizing the intra-class variation. In this case, $\mathbf{M}$ depicts the inter-class variation while $\mathbf{N}$ depicts the intra-class variation. The solutions for both problems are the eigenvectors that correspond to the top d eigenvalues of $\mathbf{M}$ for PCA and $\mathbf{N}^{-1}\mathbf{M}$ for LDA.

In two dimensional subspace techniques like 2D-PCA [14] and 2D-LDA [8], the images are retained as $h \times w$ intensity matrices instead of transforming them into a column vector. Given P training images, $\Gamma_i \in \mathbb{R}^{h \times w}$ belonging to $\mathcal{C}$ classes, the covariance matrices are computed as described in [14] and [8]. Since the images are not transformed to a column vector in the two dimensional case, the covariance matrices would $\in \mathbb{R}^{w \times w}$. Typically, in a face database $w < P \ll (h \times w)$. Due to this fact, computation of the covariance matrices using the two dimensional approach is far more efficient than the one dimensional counterpart.

2.1. Tensor Concepts

Given the importance of color in human recognition abilities, we tried to develop a scheme that incorporates color as a modality in recognition, in order to ascertain if any improvements can be achieved. One possibility of color based processing was to analyze color as a 3D tensor instead of just a vector or a matrix of values. In the field of face recognition, the earliest tensor oriented approach was by Vasilescu and Terzopoulos [13] who described a "tensorfaces" framework to handle aspects such as scene structure, illumination and viewpoint. Recently, Yu and Bennamoun [16] used these generalizations in performing face recognition using 3D face data. However, according to the authors, the biggest difficulty lay in its computational complexity.

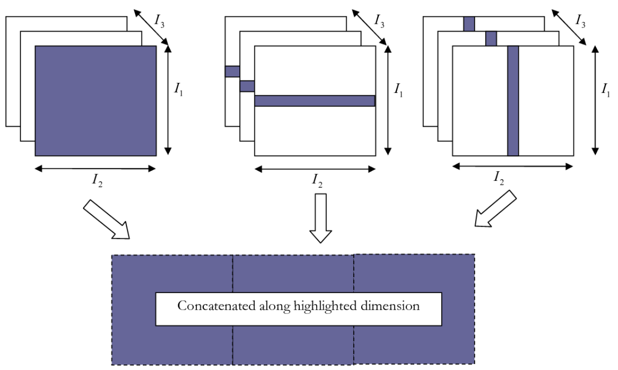


Figure 1. Typical unfolding of a 3^{rd} order tensor along the 3 possible dimensions.

For the sake of completeness, we shall briefly describe the concept of a tensor from multi-linear algebra. A tensor is a multi-linear mapping over a set of vector spaces, $\mathcal{D} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and can be seen as generalizations of vectors (1^{st} order tensor) and matrices (2^{nd} order tensor). Given a tensor $\mathcal{D}$, we can unfold it along any of its dimensions to obtain a mode-n matrix version of the tensor. For a 3^{rd} order tensor, figure 1 shows the three possible unfoldings along one of its three highlighted dimensions.

3. Algorithm

Typical face recognition algorithms are composed of two sequential steps, the face detection followed by the face recognition step. In this work, we assume that the faces have been located, either as a cropped face or as the output from an external face detection algorithm [10] and we analyze the recognition rates of our color oriented approach.

The works on two dimensional PCA [14] and two dimensional LDA [15][8] indicate an improved recognition rate over their one dimensional counterparts. One of the possible reasons for the improvement might be due to the pixel coherence that exists within any finite neighborhood when images are observed as two dimensional matrices. Typically, skin pixels would occur in close proximity to other skin pixels, except at boundaries. In the one dimensional PCA approach, this coherence is disturbed due to the vectorization of the image pixels, while the two dimensional PCA maintains the neighborhood coherence thereby achieving higher recognition accuracy.

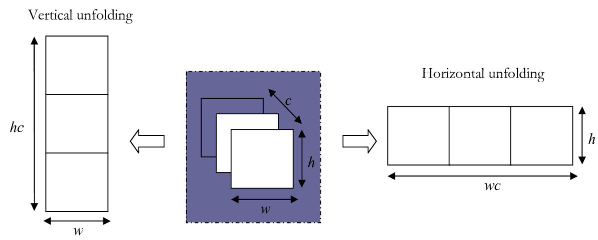


Figure 2. Horizontal and Vertical unfolding of the 3D color tensor.

When trying to tackle the color based model, we observed that the mode-1 unfolding of the 3D color tensor best maintains the coherence between local pixel neighborhoods; thus we used mode-1 unfolding for our experiments. Furthermore, mode-1 unfolding can be performed along two directions, the horizontal direction and the vertical direction. These two types of mode-1 unfolding of a $(h \times w \times c)$ image is shown in figure 2 (in our case $c = 3$).

Given P training images, we computed the vertical or horizontal mode-1 unfolding for each of the images and solved the symmetric eigenvalue problem for the two dimensional PCA and LDA respectively. In the case of a vertical unfolding, the scatter matrices $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{w \times w}$ while for horizontal unfolding $\in \mathbb{R}^{wc \times wc}$. The optimal projection axes that corresponded to the top d eigenvalues were selected as the required color face subspace.

Once the color subspace was computed, the training images were projected in the color subspace and their projection vector was stored in the face database. In the recognition stage, an input test sample was projected into the same subspace and the minimum Euclidean distance between the projected test vector and all the stored projections was computed. A nearest neighbor classification scheme was then used to identify the input test image.

4. Results and Analysis

In order to understand the variations in the different possibilities, we have tried to decompose the results based on the specific data set used to test the results. The first one was the Georgia Tech (GATech) face database collected by Ara V. Nefian[1] and the second was a pruned version of the California Institute of Technology (CalTech) face database collected by Markus Weber[2]. The GATech images were cropped png color images of 50 individuals with 15 views per individual, with no specific order in their viewing direction. The pruned Caltech database was composed of jpeg compressed color images of 19 individuals with 20 views per image. The original database was composed of 450 face images of 27 individuals, but we pruned out individuals who had lesser than 20 views to maintain uniformity within the database.

The two main differences between the images in the two databases, other than the total number of face images were the quality of face images and the percentage of face pixels in an image. The GATech database has lower resolution (quality) face images ($\sim$200$\times$ $\sim$150 pixels) and are more closely cropped with respect to the image boundary than the CalTech faces, which are at a much higher quality ($\sim$500$\times$ $\sim$350 pixels) but have more background information. Specific characteristics of the GATech database included variations in facial expressions of the individuals and the presence of spectacles among some individuals (Some views of the same individual existed with and without spectacles). The CalTech face database on the other hand had variations across facial expression along with significant illumination variation.

Due to the limitation in space, we are presenting results from the GATech face database since it represents the real world data to a better extent than the CalTech database. This is especially true if recognition is being handled using low end web cameras, where low resolution images are acquired at $\sim$30 frames per second.

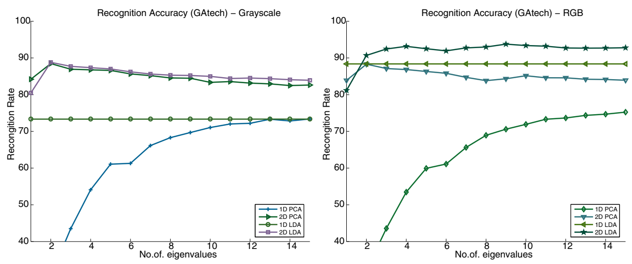
4.1. K-Fold Cross Validation

For the analysis of the face recognition accuracy, we performed a variant to the typical K-Fold cross validation [7] where K is the number of individuals in the face database. Any test configuration involved randomly selecting one view per individual and using the remaining images as the training sequence. Each configuration thus selected one view from each class of individuals for testing and used the remaining data for training.

The results from all the iterations were averaged out to compute the mean recognition rate and average subspace creation and recognition time (in seconds). For the two data sets, we have compared various constraints including 1D/2D variations, unfolding direction variations and color space variations that govern the recognition accuracy of the system.

4.2. 1D/2D Variations

Here, we tried to observe the improvement of the color space LDA approach over the traditional PCA, when we place a constraint on the "vectorization" of the image pixels. For comparative purposes, we show the variation of recognition under the grayscale assumption (the input color images were converted to gray scale) and under the assumption that all RGB components are used.



(a) (b)

Figure 3. Recognition rates for the GATech database using (a) gray scale images (b) RGB images for the 1D and 2D variations.

Figure 3(a) shows the accuracy when 1D PCA (original eigenface approach [12]), 2D PCA (approach proposed by Yang et al.[14]), 1D LDA (original fisherface approach [1]) and 2D LDA (approach proposed by Li and Yuan [8]) was applied to gray scale images. It is important to note that classical one dimensional LDA [1] has a constant recognition rate since the number of eigenvectors selected to span

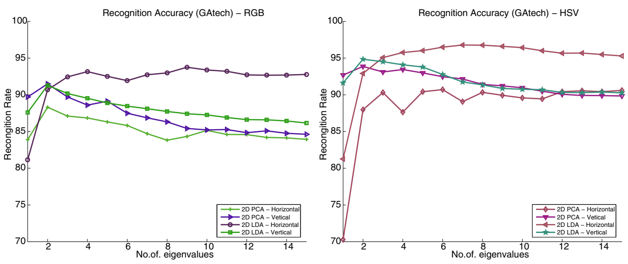
[1] `www.anefian.com/face_reco.htm`
[2] `www.vision.caltech.edu/html-files/archive.html`

the subspace are fixed to $\mathcal{C}-1$, where $\mathcal{C}$ is the number of individuals.

Figure 3(b) shows the recognition rates when RGB components are used. The color one dimensional PCA and one dimensional LDA are the original approaches extended to incorporate all the color components into the pixel space, while two dimensional PCA and two dimensional LDA are the color space oriented approaches under horizontal unfolding of the color tensor. The important observations that can be made are the improvements in one dimensional LDA and two dimensional LDA with the use of color. Between the gray scale and the color based approaches, using the color space improves recognition by around 4$\sim$5%.

4.3. Unfolding Direction Variations

In this experiment, we tested the unfolding direction and its contribution to the recognition rate. From the test cases, we observed that vertical unfolding of the 3D color tensor was better than the horizontal unfolding for the two dimensional PCA but the reverse is true for two dimensional LDA. Figures 4(a) and (b) show the results from the RGB and HSV color space where we can see that the horizontal unfolding provided a better recognition rate over the vertically unfolded version when projected into the LDA subspace.

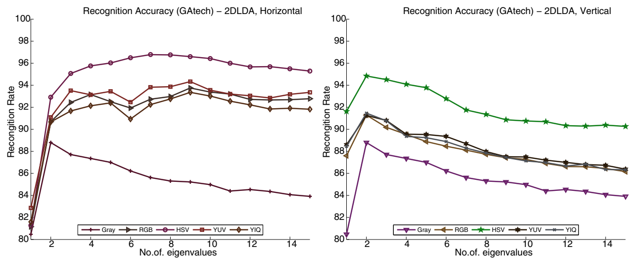


(a) (b)

Figure 4. Recognition rates for the GATech database using 2D-LDA under horizontal and vertical unfolding within (a) RGB and (b) HSV color space.

4.4. Color Space Variations

To understand the variability of the algorithms towards color space, we analyzed the recognition rate of an RGB image against 4 color spaces: Gray scale, HSV, YUV and YIQ [5]. Figures 5 (a) and (b) are the recognition accuracy of two dimensional LDA under the horizontal and vertical unfolding respectively. The conspicuous finding that can be observed from figures 5 (a) and (b) is that projection to a color subspace outperforms using a gray scale subspace. In fact, the performance of the two dimensional LDA in HSV color space shows an improvement of about 88$\sim$96% over the equivalent two dimensional LDA when applied to the gray scale color space.



(a) (b)

Figure 5. Recognition rates for for the GATech database using 2D-LDA under different color spaces for (a) horizontal and (b) vertical unfolding.

4.5. Efficiency Analysis

For computing the time required to perform the operations, the entire processing was broken down into three time segments. The first segment was the time needed to create the scatter matrices ($\mathbf{M}$ for PCA or $\mathbf{M}$ and $\mathbf{N}$ for LDA). The second segment was the time to project the training data into the subspace and finally, the time required to recognize the test images.

Among the three segments, the first part was typically the one which required the most amount of time but this is typically a one time cost and thus can be amortized over an entire session. The second and third segments varied depending on the dimension of the subspace that were used for projecting the training and testing faces. The system was implemented in Matlab and we believe that the computational improvement would be proportional if the development was in "C".

5. Conclusions

In this paper, we have shown that the utilization of color cues could improve the accuracy of the face recognition algorithm by 8$\sim$10%. We computed a mode-1 unfolding of a color tensor to maintain maximum neighborhood coherence and applied two dimensional LDA on the unfolded color tensor. From the quantitative experiments, we have observed that the color space oriented approach improved recognition accuracy when compared to the gray scale approach. Differences between the various color spaces have not been conclusively established and further research is needed to see which color space performs better. Another important observation is that horizontal unfolding of the color tensor improves recognition rate as against vertical unfolding for the two dimensional LDA while the reverse holds for the two dimensional PCA. This is consistently observed irrespective of the color space, and we are currently trying to understand the reason for these differences in the unfolding direction.

References

[1] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720, 1997.

[2] T. D. Bie, N. Cristianini, and R. Rosipal. *Handbook of Geometric Computing*, chapter Eigenproblems in Pattern Recognition, pages 129–167. Springer Berlin Heidelberg, 2005.

[3] J. R. del Solar and P. Navarrete. Eigenspace-based face recognition: a comparative study of different approaches. *IEEE Trans. Sys., Man and Cyber. - Part C: Applications and Reviews*, 35(3):315–325, 2005.

[4] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, 15(5):252–260, 2005.

[5] A. Ford and A. Roberts. Colour space conversions. August 1998.

[6] B. Heisele, T. Serre, and T. Poggio. A component-based framework for face detection and identification. *International Journal of Computer Vision*, 74(2):167–181, August 2007.

[7] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proc. of the 14th International Joint Conference on Artificial Intelligence*, pages 1137–1145, 1995.

[8] M. Li and B. Yuan. 2D-LDA: A statistical Linear Discriminant Analysis for image matrix. *Pattern Recognition Letters*, 26:527–532, 2005.

[9] Q. Li, J. Ye, M. Li, and C. Kambhamettu. Adaptive appearance based face recognition. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI '06)*, pages 677–684, 2006.

[10] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *Proc. of the International Conference on Image Processing*, pages 900–903, September 2002.

[11] G. Shakhnarovich and B. Moghaddam. *Handbook of Face Recognition*, chapter Face Recognition in Subspaces. Springer-Verlag, 2004.

[12] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[13] M. A. O. Vasilescu and D. Terzopoulos. Multilinear image analysis for facial recognition. In *Proc. of the International Conference of Pattern Recognition (ICPR 2002)*, volume 2, pages 511–514, 2002.

[14] J. Yang, D. Zhang, A. F. Frangi, and J. Yang. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):131–137, 2004.

[15] J. Ye, R. Janardan, and Q. Li. Two-dimensional linear discriminant analysis. In *Advances in Neural Information Processing Systems*, pages 1569–1576, 2005.

[16] H. Yu and M. Bennamoun. 1D-PCA, 2D-PCA to nD-PCA. In *Proc. of the International Conference of Pattern Recognition (ICPR2006)*, volume IV, pages 181–184, 2006.

[17] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.

Method	(C, D)	E	RR	S	T_1	T_2
1D-PCA	Gray, -	15	73.38	9.75	0.092	0.034
1D-PCA	RGB, -	15	75.24	21.34	0.26	0.064
1D-PCA	HSV, -	15	**82.27**	21.6	0.26	0.064
1D-PCA	YIQ, -	15	75.0	21.3	0.26	0.067
1D-PCA	YUV, -	15	75.47	21.34	0.26	0.065
1D-LDA	Gray, -	49	73.33	16.65	0.019	0.39
1D-LDA	RGB, -	49	88.38	35.55	0.019	1.05
1D-LDA	HSV, -	49	**92.53**	35.85	0.019	1.05
1D-LDA	YIQ, -	49	89.6	35.55	0.019	1.05
1D-LDA	YUV, -	49	89.1	35.54	0.019	1.05
2D-PCA	Gray, -	2	88.47	0.82	0.07	0.34
2D-PCA	RGB, $\mathcal{H}$	2	88.3	5.77	0.23	0.38
2D-PCA	HSV, $\mathcal{H}$	6	**90.7**	5.77	0.3	1.3
2D-PCA	YIQ, $\mathcal{H}$	2	88.7	5.63	0.24	0.38
2D-PCA	YUV, $\mathcal{H}$	2	88.9	5.66	0.23	0.38
2D-PCA	RGB, $\mathcal{V}$	2	91.49	2.53	0.22	1.29
2D-PCA	HSV, $\mathcal{V}$	2	**93.87**	2.53	0.47	9.93
2D-PCA	YIQ, $\mathcal{V}$	2	91.7	2.51	0.21	1.29
2D-PCA	YUV, $\mathcal{V}$	2	91.76	2.54	0.21	1.29
2D-LDA	Gray, -	2	88.8	1.23	0.079	0.35
2D-LDA	RGB, $\mathcal{H}$	9	93.76	6.8	0.26	2.01
2D-LDA	HSV, $\mathcal{H}$	7	**96.78**	6.99	0.2	1.54
2D-LDA	YIQ, $\mathcal{H}$	9	93.35	6.93	0.34	2.01
2D-LDA	YUV, $\mathcal{H}$	9	94.33	7.23	0.34	2.02
2D-LDA	RGB, $\mathcal{V}$	2	91.29	3.65	0.21	1.31
2D-LDA	HSV, $\mathcal{V}$	2	**94.85**	3.62	0.22	1.31
2D-LDA	YIQ, $\mathcal{V}$	2	91.47	3.62	0.22	1.31
2D-LDA	YUV, $\mathcal{V}$	2	91.27	3.66	0.21	1.31

Table 1. Computational time (in seconds) for the GATech face database (C - Color space, D - Unfolding direction, RR - best recognition accuracy, E - subspace dimension for best recognition accuracy, S - Subspace creation time, T_1 - time to project training sequence, T_2 - time to recognize test sequence, $\mathcal{H}$ - Horizontal unfolding, $\mathcal{V}$ - vertical unfolding).