Enhanced Video Indexing and Retrieval Based on Face Recognition through Combined Detection and Fast LDA

Loganathan D., Jamal J., Nijanthan P., and Kalichy Balamurugan V.

Department of Computer Science and Engineering,
Pondicherry Engineering College, Puducherry - 605014, India
{drloganathan,jamal,nijanth.niji}@pec.edu,
kalichy.90@gmail.com

Abstract. The content based indexing and retrieval of videos plays a key role in helping the Internet today to move towards semantic web. The exponential growth of multimedia data has increased the demand for video search based on the query image rather than the traditional text annotation. The best possible method to index most videos is by the people featured in the video. The paper proposes combined face detection approach with high detection efficiency and low computational complexity. The fast LDA method proposed performs wavelet decomposition as a pre-processing stage over the face image. The preprocessing stage introduced reduces the retrieval time by a factor of $1/4^n$ where n is the level of decomposition as well as improving the face recognition rate. Experimental results demonstrate the effectiveness of the proposed method reducing the retrieval time by 64 times over the direct LDA implementation.

Keywords: Video Indexing and Retrieval, Face Detection, Face Recognition, Fast LDA, Wavelet Transformation.

1 Introduction

Digital image and video are rapidly evolving as the modus operandi for information creation, exchange and storage in modern era over the Internet. The videos over the Internet are traditionally annotated with keywords manually. The fast growth of videos over the past few decades has increased the demand of a query by example (QbE) retrieval system in which the retrieval is based on the content of the videos [1].

Face detection and recognition techniques besides being used extensively in authentication and identification of users, has also been extended to index and retrieve videos [2]. People are the most important subjects in a video. Face detection is used to identify faces in the image sequences and face recognition is used to associate the video with the people featured in the video. The face recognition algorithms classified into two types namely appearance based and geometric feature based approach [3]. The latter systems are computationally expensive compared to the former [4].

Wavelet transform has been a prominent tool for multi-resolution analysis of images in the recent past [5]. 2D Discrete Wavelet Transform (DWT) decomposes the

V.V. Das and J. Stephen (Eds.): CNC 2012, LNICST 108, pp. 351–357, 2012.

image into four components namely Approximation (cA) corresponding to low frequency variations, Horizontal (cH) corresponding to horizontal edges, Vertical (cV) corresponding to vertical edges and Diagonal (cD) corresponding to non-horizontal and non-vertical edge components as in fig. 4.

This paper proposes a system that uses a combination of skin color models to identify the skin regions, followed by morphological and geometric operations to find the face regions. The face image to recognition phase is pre-processed by wavelet decomposition and the approximation component is fed as input, thereby increasing recognition rate and reducing time complexity by nearly 64 times.

2 Framework

The overall architecture as in Fig. 1 is based on the system proposed in [2]. Video data can be viewed as a hierarchy, made up of frames at the lowest level. A collection of frames focusing on one object depicts a shot. [6]. The key frames extracted after shot detection [7] [8] are then subjected to the Combined Face Detection method. The face images obtained are preprocessed by wavelet decomposition. In the retrieval phase, the images in the training database are used both for training (building the projection space) and testing stage (identifying a close match to test image).

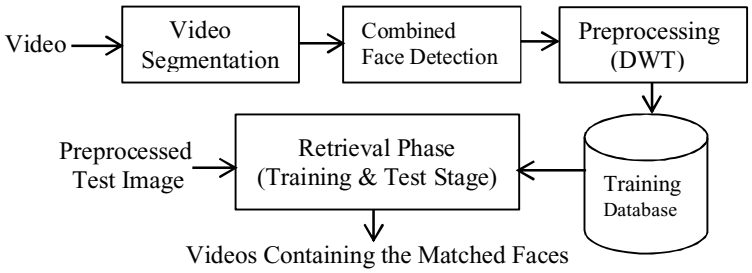

Fig. 1. Framework of the Video Indexing and Retrieval System

3 Combined Face Detection

The face detection method proposed in this paper tries to maximize the detection efficiency at the same time reducing the computational complexity.

3.1 Combined Skin Color Models

The image is converted into the three color spaces namely Normalized RGB, YCbCr and HSV color model [12]. The threshold values for skin detection in each color space given in Table 1 are applied to obtain binary images as shown in fig. 2.

Table 1 Threshold Values in each Color Space for skin Detection

Color Model	*Threshold Values*
Normalized RGB	$0.41 < r < 0.50$ & $0.21 < g < 0.30$
YCbCr	$133 < Cr < 173$ & $77 < Cb < 127$
HSV	$6 < H < 38$

The unions of all binary images are shown in Fig. 3a. This step is intuitive of the basic set theory principles. Thus if a region is detected as skin region in any one of the color spaces is recognized as a skin region. This overhead of converting the image into three color spaces in lieu of one is circumvented by the added advantage of reduced false rejection rate.

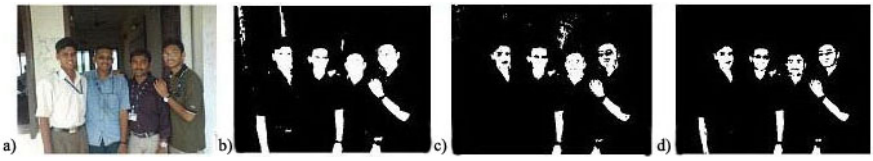

Fig. 2. Results of image after applying threshold on a)original image b) Normalized RGB c) YCbCr d) HSV color models

3.2 Morphological Operations

Before application of morphological processing, the Euler number (the total number of objects minus the number of holes in those objects) of the binary image is calculated. Flat regions having no holes such as exposed hand, leg or a background having the same color as skin region may be excluded from further processing. Morphological opening and closing are applied to the resultant image. Then fill operation are applied to remove holes and form connected regions as shown in fig. 3b.

3.3 Geometrical Analysis of Connected Regions

The connected regions are analyzed to check if it conforms to the geometry of a face region. The face regions cannot be too narrow, too short, short and wide or narrow and tall and are excluded. The ratio of height to width of the remaining regions are calculated and checked if it conforms to the Golden Ratio, S given by (1). Fig. 3c is the result of ANDing the mask obtained with the original image.

$$S = \frac{1 + \sqrt{5}}{2}. \tag{1}$$

3.4 Face Region Identification

The centroids of each of the connected region are calculated and an elliptical mask around the face region is drawn. The masks created are then AND with the original frame to obtain the face regions in the frame as shown in fig. 3d.

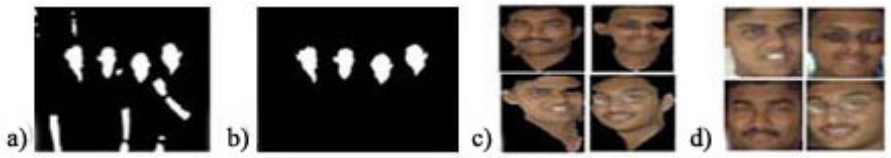

Fig. 3. a) Union of binary images b) Resultant after Morphological & Geometrical Analysis c) Selected regions by ANDing d) Cropped Images

4 Fast Linear Discriminant Analysis

The Fast LDA involves the wavelet decomposition by Daubechies wavelet 'db4' of input image. The approximation component having high intensity variation and recognition rate over other components [4] are then subjected to LDA.

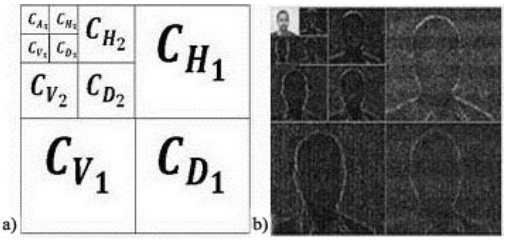

Fig. 4. a) Co-efficient of wavelet decomposition b) 'db4' at level 3 applied to image

Linear subspace methods suffer from large computational load which arises due to the large dimensions of the Eigen subspace. The approximation component ($\boldsymbol{C_{A_3}}$) is used as input to the LDA. This component with reduced dimension retains the high intensity variance. This pre-processing stage though has the overhead of wavelet decomposition to three levels, reduces the recognition time by a factor of $\boldsymbol{1/4^n}$, n is the depth of decomposition, owing to the reduced dimension of the Fisher face space.

Fig. 5. Preprocessed images of sample faces from ORL database

LDA maintains class information where a class is a set of images of a person in different poses. Let N be the total number of images, N_i be the number of images in each class and C be the total number of classes. Let m be the mean image across all classes and m_i be the mean image of each class. The pre-processed image is reshaped as column vector x_j of dimension $lm\ x\ 1$. The within class scatter matrix is given by

$$S_w = \sum_{i=1}^{C} \sum_{j=1}^{N_i} (x_j - m_i)(x_j - m_i)^T . \quad (2)$$

The between class scatter matrix can be considered as the variance of each class across all images and is given by

$$S_b = \sum_{i=1}^{C} (m_i - m)(m_i - m)^T . \tag{3}$$

The main objective of LDA is to maximize the ratio between S_b and S_w, known as the Fisher's criterion given in (4). The problem reduces to solving the generalized Eigen equation in (5).

$$\arg\max(w) \frac{|W^T S_b W|}{|W^T S_w W|} . \tag{4}$$

$$S_b W = S_w W \lambda . \tag{5}$$

where λ is a diagonal matrix containing the Eigen values of ${S_w}^{-1} S_b$ and W is the corresponding Eigen vectors. The S_w happens to be singular in almost all cases for the number of training images should be almost equal to size of the scatter matrices. The singularity problem is solved by using the Direct Fractional Step LDA [9].

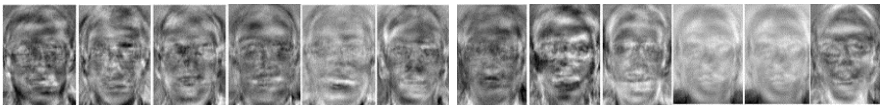

Fig. 6. Sample Fisher faces applying LDA on ORL Database

Once the transformation matrix W is calculated, the projection of the mean image of each class on the LDA subspace is calculated by

$$Y_j = W^T m_j . \tag{6}$$

$$Y = [\, Y_1 Y_2 Y_3 \ \ldots \ Y_C \,] . \tag{7}$$

Fig. 7. Sample Fisher faces of fast LDA on ORL Database

The highest Fisher faces obtained by applying LDA and fast LDA are shown in fig. 6 and 7. The proposed algorithm projects mean image of each class on to the projection space rather than all the images in a class as in (6). Let the test image vector be x_t. The test image is projected on to the subspace given by

$$Y_t = W^T x_t . \tag{8}$$

The Euclidean distance between the test image and each class is calculated by

$$\delta_j = \left|\left| Y_t - Y_j \right|\right| = \sqrt{\sum_{j=1}^{C} (Y_t - Y_j)^2} . \tag{9}$$

The class which has the least Euclidean measure with respect to the test image is considered the match and associated videos are retrieved.

5 Experimental Results

The face detection and recognition system has been tested with different video streams. The video types include news, talk shows having transition effects such as fade, abrupt transition, wipe and dissolve. The proposed system focuses on improving the retrieval time to facilitate recognition in real time and for large database. The results presented in this section conforms to the implementation of the algorithms proposed and the system as whole in MATLAB r2010 on Intel Core 2 Duo Processor running Windows 7.

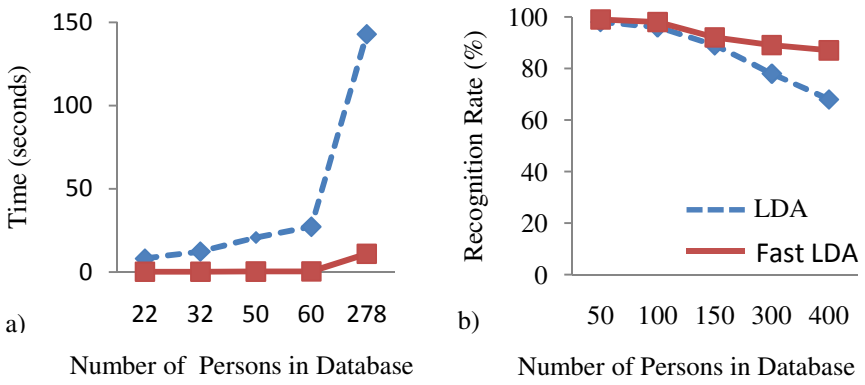

Fig. 8. Performance of the proposed fast LDA algorithm against the LDA in a) Recognition Time b) Recognition Rate

Fig. 8 shows the performance of LDA and the fast LDA with respect to the retrieval phase over the face databases namely ORL database, Indian Face database [10], and MUCT database [11]. An integrated system for video indexing and retrieval is built with proposed enhancement. The MPEG video sequences having 30 frames/second are considered. Table 2 gives the details of the comparative performance of the system.

Table 1. Results of the proposed algorithm to video database

Number of Videos (Avg. Length)	*Avg. Recognition time in LDA (sec)*	*Avg. Recognition time in fast LDA (sec)*
20 (12 min)	80.23	1.28
40 (15 min)	120.82	5.89

6 Conclusion and Future Work

The papers proposes the combined face detection and fast LDA methods which improves the recognition rate and reduces the retrieval of videos based on face recognition making the system suitable for large databases. Further work in the same line includes analyzing methods for faster implementation of wavelet decomposition to reduce the extra overhead in the indexing phase.

References

1. Hu, W., Xie, N., Li, L., Zeng, X.: A Survey of Visual Content Video Indexing and Retrieval. J. IEEE 41(6), 797–819 (2011)
2. Torres, L., Vila, J.: Automatic Face Recognition for Video Indexing Applications. J. Pattern Recognition 35(3), 615–625 (2002)
3. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and Machine Recognition of Faces: a Survey. IEEE 83(5), 705–741 (1995)
4. Chellpa, J., Etemad, K.: Discriminant Analysis for Recognition of Human Face Images. J. Optical Society of America 14(8), 1724–1733 (1997)
5. Todd Ogden, R.: Essential Wavelets for Statistical Applications and Data Analysis. Birkhäuser, Boston (1997)
6. Monaco, J.: How to Read a Film: The Art, Technology, Language, History, and Theory of Film and Media. Oxford University Press (1977)
7. Yusoff, Y., Christmas, W., Kitter, J.: Video Shot Cut Detection Using Adaptive Thresholding. In: British Machine Vision Conference (2000)
8. Boreczsky, J.S., Rowe, L.A.: Comparison of Video Shot Boundary Detection techniques. In: SPIE Conference on Video Database, pp. 170–179 (1996)
9. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face Recognition Using LDA Based Algorithms. IEEE 14(1), 195–200 (2003)
10. The Indian Face Database (2002), http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/
11. Milborrow, S., Morkel, J., Nicolls, F.: MUCT database. University of Capetown (2008)
12. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing Using MATLAB. Tata McGraw Hill, New Delhi (2011)