International Journal For Technological Research In Engineering
Volume 1, Issue 11, July-2014 *ISSN (Online): 2347 - 4718*

FACE DEPENDENT VIDEO SUMMARIZATION

Namitha A R[1], Deepu R[2], Honnaraju B[3]
[1]M. Tech, [2]Asso Prof & HOD, [3]Asso Prof
Department of Computer Science
Maharaja Institute of Technology, Mysore, India

Abstract: Video over Internet is getting more popular nowadays, since the rapid development of the Internet technologies, with the higher network bandwidth, people can retrieve information in the form of multimedia including image, audio, and particularly video. Face detection is essentially localizing and extracting a face region from the background. Video summary is a process of presenting an abstract of entire video within a shorter period of time. It also aims to provide a condensed video representation, while preserving the sequence of shots, as of the original video. Face recognition is a process of recognizing the face in a frame by comparison among the samples. Consider video sample is retrieved and divided into frames, then faces are detected in the video frames and stored in a separate folder. Using PCA, features will be extracted in the frames where faces are detected in video frames, then the test images are stored at the database in prior, to train the system for comparison. In this paper the PCA features for feature extraction are used and matching is done for under consideration with the test image using Eigen face coefficients.

I. INTRODUCTION

A. Overview

Digital video processing has been extremely active over the past decades. The importance of this area from the explosion of video resources from the internet, e.g.google and yahoo video , youtube etc. Shot Detection in a video sequence is a process of identifying visual discontinuities along the time domain. During this process, it is required to extract visual features that measure the degree of similarity between frames in a given shot. Video summary is a process of presenting an abstract of entire video within a shorter period of time. It also aims to provide a condensed video representation, while preserving the sequence of shots, as of the original video. It can also allow us to extract the required information when we are not interested in the whole unit. It help us to assess the relevance or value of information within a shorter period of time. The face plays a major role in our social intercourse in conveying identity and emotion. The human ability to recognize faces is remarkable. We can recognize thousands of faces learned throughout our lifetime and identify familiar faces at a glance even after years of separation. The skill is quite robust, despite large changes in the visual stimulus due to viewing conditions, expression, aging, and distractions such as glasses or changes in hairstyle. The face of the person are the first information looked for an image. Performing efficient face detection in a video with persons (excluding cartoons and nature videos) allow to classify shots, and to obtain automatically face summaries. Shot sampling greatly improves time processing. Scene layout (same number of person, similar face position and size) provides a criterion to establish a similarity measure between shots. Person are the principal shooting operators, the major subject of shooting and also the primary concern of audience. This concern and aptitude for persons and specifically for person faces is illustrated every day in television magazines (paper or electronic) where, it is a well-established convention, summaries of programs are quasi systematically illustrated with images of persons and most of the time with (cropped) close-up shots of face. An alternative to summarizing a whole video to a unique image consists in segmenting the video in shots. Segmentation consists in finding the location and the nature of the transition between two adjacent shots and has led to numerous techniques, trimmed to the nature of the transition both in the compressed [1] ,uncompressed domain[2]. Each identified shot is summarized by a keyframe [3]. Shot detection and key frame extraction rely on low level information (colour, movement) but nothing is known on the content of the shot or key frame (presence/absence of persons or of specific objects). They present a technique to summarize video using face information obtained by face detection. This technique is adequate for videos with persons but unsuitable for videos such as cartoons or nature documentaries (without faces of persons). Face recognition is challenging because it is a real world problem. The human face is a complex, natural object that tends not to have easily (automatically) identified edges and features. Because of this, it is difficult to develop a mathematical model of the face that can be used as prior knowledge when analyzing a particular image. Computational models of faces have been an active area of research since late 1980s, for they can contribute not only to theoretical insights but also to practical applications, such as criminal identification, security systems, image and film processing, and human-computer interaction, etc. However, developing a computational model of face recognition is quite difficult, because faces are complex, multidimensional, and subject to change over time. The human capacity to recognize particular individuals solely by observing the human face is quite remarkable. This capacity persists even through the passage of time, changes in appearance and partial occlusion. Because of this remarkable ability to generate near-perfect positive identifications, considerable attention has been paid to methods by which effective face recognition can be replicated on an electronic level. Certainly, if such a complicated process as the identification of a human individual based on a method as non-invasive as

face recognition could be electronically achieved then fields such as bank and airport security could be vastly improved, identity theft could be further reduced and private sector security could be enhanced. Many approaches to the overall face recognition problem (The Recognition Problem) have been devised over the years, but one of the most accurate and fastest ways to identify faces is to use what is called the "eigenface" technique. Face Recognition is a field of multidimensional applications. A lot of work has been done, extensively on the most of details related to face recognition. This idea of face recognition using PCA is one of them. In this paper the PCA features for Feature extraction are used and matching is done for the face under consideration with the test image using Eigen face coefficients. The Principal Component Analysis [4] is one of the of the most powerful techniques that have been used in image recognition or in compression. PCA is a statistical method under the broad title of factor analysis. The function of PCA is to reduce the large size of the data space (variables) to the smaller intrinsic dimensionality or size of feature space (independent variables), that are needed to describe the data cost efficiently. This is the case when there is a strong correlation between observed variables. In various functions of PCA are discussed. Because PCA is a classical technique which can perform functions in the linear domain, thus the applications having linear models are much suitable. The field of Face recognition has so many areas of application as in security, biometric systems, banks and many more that are beyond the list. Moreover, face recognition can be partitioned into Face identification, Face classification, sex determination, people surveillance in crowded areas, Video content indexing, Personal identification (e.g. Driver's License), Mug shots matching and Entrance security. The main idea of using PCA for face recognition is to express the large 1-D vector of pixels constructed from 2-Dfacial image into the compact principal components of the feature space. This can be called projection of eigen space. Eigen space is calculated by identifying the eigenvectors of the covariance matrix derived from a set of facial images(vectors). Once the eigenfaces have been computed, several types of decision can be made depending on the application. Face recognition is a broad term which is categorized as identification where the labels of individuals must be obtained, categorization where the face must be assigned to a certain class. Recognition of a person, where it must be decided if the individual has already been seen, PCA computes the basis of a space which is represented by its training vectors. These basis vectors, actually eigenvectors, computed by PCA are in the direction of the largest variance of the training vectors called eigenfaces. Each eigenface[5] can be viewed a feature. When a particular face is projected onto the face space, its vector into the face space describes the importance of each of those features in the face. The face is expressed in the face space by its eigenface coefficients. We can handle a large input vector, facial image, only by taking its small weight vector in the face space. This means that we can reconstruct the original face with some error, since the dimensionality of the image space is much larger than that of face space. Each face in the training set is transformed into the face space and its components are stored in memory. The face space has to be populated with these known faces. An input face is given to the system, and then it is projected onto the face space. The system computes its distance from all the stored faces.

II. METHODOLOGY

In this section, we describe the algorithm we used in the face detection and face recognition. After video processing only the frames which have faces are detected and rest of frames are discarded. In order to identify the frames consisting of faces viola jones algorithm is used. The basic principle of the Viola-Jones algorithm is to scan a sub-window capable of detecting faces across a given input image. The standard image processing approach would be to rescale the input image to different sizes and then run the fixed size detector through these images. This approach turns out to be rather time consuming due to the calculation of the different size images. Contrary to the standard approach Viola-Jones rescale the detector instead of the input image and run the detector many times through the image – each time with a different size. Eigenface method uses PCA for reduction in the dimension; it yields projection directions which maximize the total scatter across all images of the faces. In doing so it retains the unwanted variations due to changes in illumination. This in turn causes errors when it comes to discrimination between faces. However it is mentioned [4] that discarding the three most significant principal components help in reducing the effects caused by illumination assuming that these variations were in the first place captured by these components.

III. EIGENFACES ALGORITHM

The process is based on the proposed method by Turk and Pentland to initially train the system were

- The images of all the different individuals were collected in a set called the training set.
- The Eigenfaces were calculated from the training set but only M Eigenvectors corresponding to M largest eigenvalues were retained. These comprised the face space of the entire image database.
- Next each image in the training set is projected onto this image space and corresponding weights are calculated. Now each image is represented by M weights. Thus the entire database is very compactly represented.
- The test image now undergoes the same process and is now represented by M weights.
- The next step is to compute the distance of the M weights to the M weights of each image in the training set.

The minimum of these distances is calculated and the image from the database is chosen as the match or the closest match with which it has this minimum distance, for calculating minimum distance we are using eucliden distance.

IV. EXPERIMENTAL RESULTS

Firstly, we collect video of different types like Youtube, Google and Yahoo video, AOL videos etc. Here we are using avi videos for good accuracy purpose. Videos may be news video, classroom video, anchor videos. After some preprocessing is required to enhance specific image features. Image enhancement is performed to improve the raw image by suppressing noise and to emphasis structures. We have to separate the frames from video stream. This process continues until the video ends.

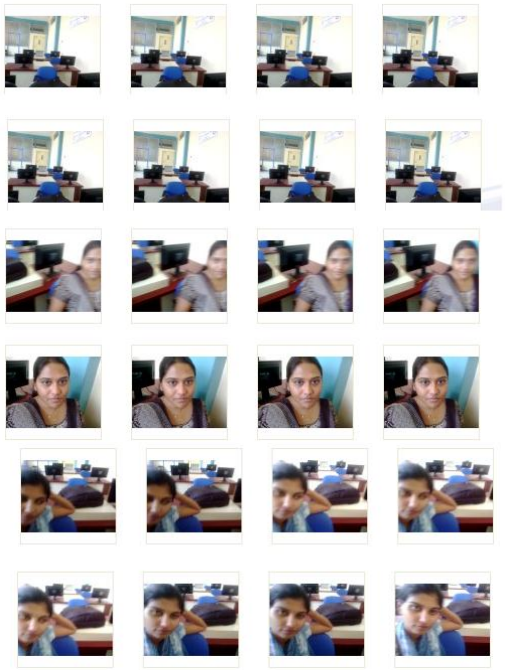

Fig 1 Shows extract frames from classroom.avi video

After video processing only the frames which have faces are detected and rest of frames are discarded. Inorder to identify the frames consisting of faces viola jones algorithm is used. The bounding box is applied to detect the face in the identified frames.

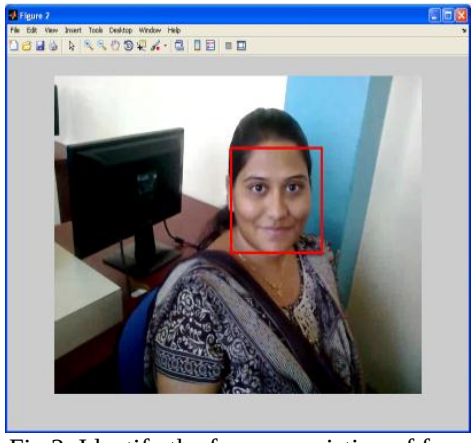

Fig 2: Identify the frame consisting of face

Later cropped image is displayed as a separate image which will be stored in the folder.

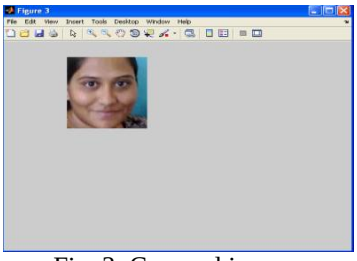

Fig. 3: Cropped images

Fig 4: Video summarization of faces

After obtain video summarization of faces(Fig 3.5)and collecting different types of videos summarization face expressions(Fig.3.6).One out of the several images is taken as the test image(Fig.3.7) and other variations of the same image is taken as train image(Fig.3.8). A separate data base is created for both train and test images.

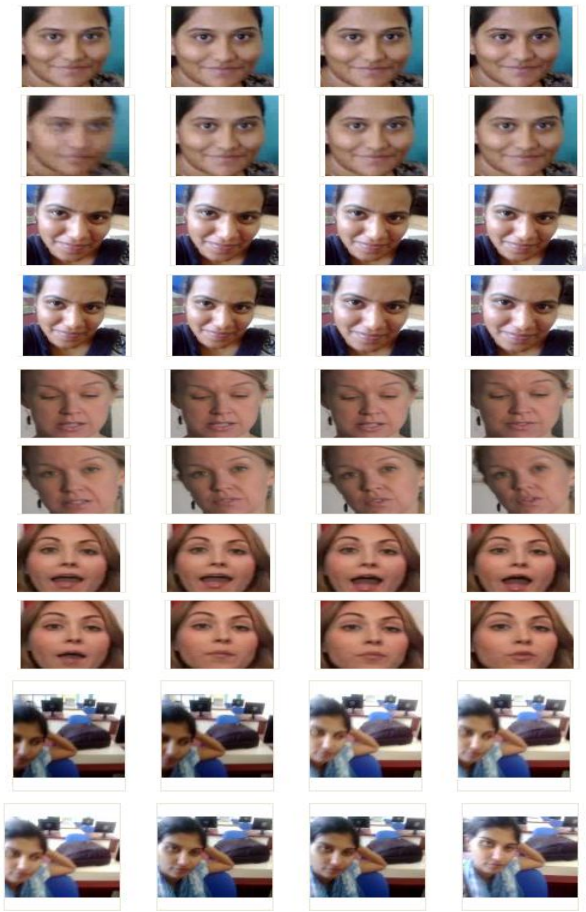

Fig.5: Different types of video summarization of faces expression

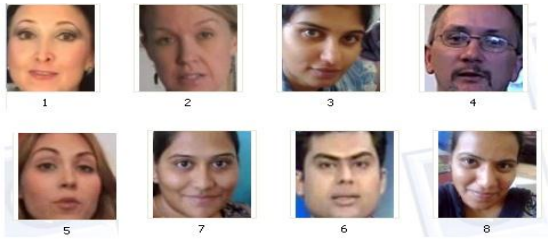

Fig.6: Creation of test images database

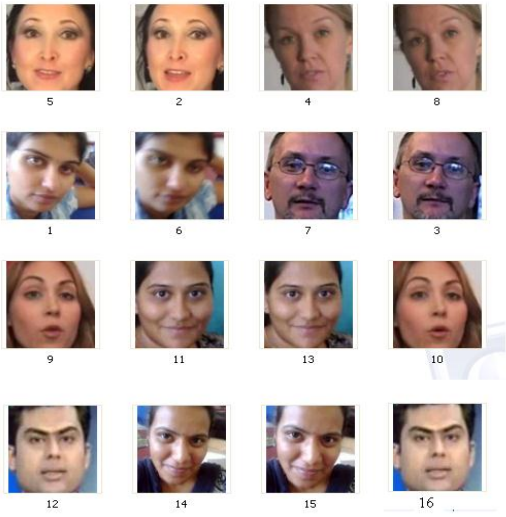

Fig.7:Creation of train images database

After the test images are stored at the database in prior, to train the system for comparison then recognized image shown in figure 3.9.

Recognized image: 13.jpg

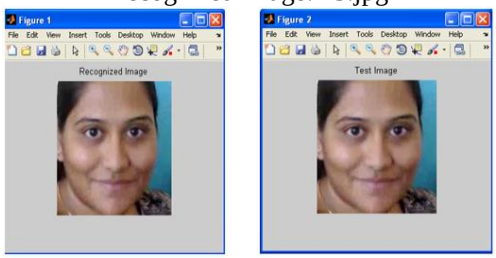

Fig.8**:** Recognized Image

V. COMPARISON OF FACE RECOGNITION METHODS

The Euclidian distance classifier is an efficient classification technique in areas where clusters of points representing the different entities to be classified are spaced far apart in feature space. The idea behind the Euclidian distance classifier is that one computes the average of several training vectors for each possible categorization and then classifies a given test vector by determining to which cluster the average is the vector nearest. The neural network classifier is a more advanced alternative to the Euclidian distance method. The primary advantage of the neural network is that no assumptions need to be made as to how clustering will occur in face-space. Given the correct network configuration, situations like the tongue out versus normal pose problem should be handled automatically by the system, with no additional design consideration. The primary conclusion that can be drawn from the results of the experiments is that the Euclidian distance classifier seems to have many advantages over its neural network counterpart. The Euclidian distance classifier matched or outperformed the neural network classifier in accuracy for every test. It is surprising that it was not able to do better on the other tests. Perhaps even more importantly the Euclidian distance classifier was faster than the neural network in both training and classifying. For most of the experiments, the Euclidian distance classifier could be trained within a matter of seconds, while the training of the neural network took anywhere from several minutes to an hours, depending on the initial random weights created in the network [6].

VI. CONCLUSION

Face detection is a mean to obtain video summaries, which people are familiar with that to focus on face information. The size of the obtained video summaries is far smaller than the standard shot summary and even benefits from non-detected faces together with a low false alarm rate. Many of the face images are similar and can be gathered in shot clusters and discarded from the summary. An eigenfaces approach uses Principal Component Analysis (PCA) for the recognition of the images. This method represents a face by projecting original images onto a low-dimensional linear subspace—'face space', defined by eigenfaces. A new face is compared to known face classes by computing the distance between their projections onto face space. One of the major advantages of eigenfaces recognition approach is the ease of implementation. Furthermore, no knowledge of geometry or specific feature of the face is required; and only a small amount of work is needed regarding preprocessing for any type of face images.

REFERENCES

[1] Wang, H.L,Chang, S.F.:"A Highly Efficient System for Automatic Face Region Detection in MPEG Video". CirSys Video, 7(4) (1997) 615-628.

[2] Demarty,C.H ,Beucher,S.:"Efficient morphological algorithms for video indexing", Content-Based and Multimedia Indexing, CBMI'99 (1999) .

[3] Chen, J.-Y., Taskiran C., Albiol, A., Delp, E. J., Bouman, C. A.: ViBE: "A Video Indexing and Browsing Environment", Proceedings of the SPIE Conference on Multimedia Storage and Archiving Systems IV, 20-22 septembre, Boston, vol. 3846 (1999) 148-164 .

[4] Kyungnam Kim, "Face Recognition using Principle Component Analysis", International Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1996.

[5] Wendy S.Yambor, "Analysis of PCA Based and Fisher Discriminant-Based Image Recognition Algorithms", M.S. Thesis, July 2000 (Technical Report CS-00-103, Computer Science).

[6] Christopher Burns and Ryan Morlok, "A Comparison of Face Recognition Methods"May 8, 2003.

