Middle-East Journal of Scientific Research 20 (4): 537-541, 2014
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2014.20.04.21040

Video Key Frame Extraction for Recognising Hand Drawn Human Face

[1]T. Judes Divya and [2]A. Rama

[1]Department of Information and Technology, India
[2]Tagore Engineering College, Bharath University, India

Abstract: Face Recognition is one of the biometric methods identifying individuals by the features of face. An important application of face recognition is to assist law enforcement. This paper presents the video key frame extraction for face recognition in video. The video sequences are segmented into shots by detecting gradual and abrupt cuts. Further longer shots are segmented into sub shots based on location and camera motion features. Each shot and sub shot contains a number of frames. The frames that contain the human faces are extracted as key frames. Since face sketches represent the original faces in a very concise yet recognizable form, they play an important role in criminal investigations, human visual perception and face biometrics. A hand drawn face sketch is given as a query image. The query image is compared with the faces in the extracted key frame. For the comparison of query image and faces in the key frame, Principle Component Analysis (PCA) algorithm is used and thus face recognition is performed in video.

Key words: Key frame extraction · Principle Component Analysis · Hand drawn Sketch

INTRODUCTION

Facial Recognition system is a computer technology that determines the locations and sizes of human faces in arbitrary (digital) images. It detects facial features and ignores anything else, such as buildings, trees and bodies. A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. One of the ways to do this is by comparing selected facial features from the image and a facial database. Face detection is the first stage of an automatic face recognition system, since a face has to be located in the input image before it is recognized. A definition of face detection could be: given an image, detect all faces in it (if any) and locate their exact positions and size [1]. Some facial recognition algorithms identify faces by extracting landmarks, or features, from an image of the subject's face. For example, an algorithm may analyze the relative position, size and/or shape of the eyes, nose, cheekbones and jaw. These features are then used to search for other images with matching features.

Popular recognition algorithms include Principal Component Analysis with eigenface, Linear Discriminate Analysis, Elastic Bunch Graph Matching fisher face, the Hidden Markov model and the neuronal motivated dynamic link matching. Facial recognition system is gaining the interest of marketers. This face processing system finds numerous applications in face tracking, facial expression recognition, facial feature extraction, gender classification, clustering, attentive user interfaces, digital cosmetics, biometric systems. In addition, most of the face detection algorithms can be extended to recognize other objects such as cars, humans, pedestrians and signs, etc [2].

Facial recognition is most beneficial to use for facial authentication than for identification purposes, as it is too easy for someone to alter their face, features with a disguise or mask, etc. Depending on the application, a face recognition system can be working either on identification or verification mode. Facial recognition, when used in combination with another biometric method, can improve verification and identification results dramatically.

Facial recognition in videos has become an interesting research area. This system is highly useful for identifying criminal activities and criminals from the surveillance videos. Facial recognition system in videos is the facial recognition system along with key frame extraction. For video, a common first step is to segment the videos into temporal "shots," each representing an event or continuous sequence of actions. A shot

Corresponding Author: T. Judes Divya, Department of Information and Technology, India.

represents a sequence of frames captured from a unique and continuous record from a camera. Then key frames are to be extracted. Video segmentation is the premise of key frame extraction and key frames are the salient content of the video.

Related Work: There is a large body of literature on face recognition. Performance of the DF-LDA method is overall superior to those of traditional FR approaches, such as the Eigen faces, Fisher faces and D-LDA methods. The new method utilizes a new variant of D-LDA to safely remove the null space of the between-class scatter matrix and applies a fractional step LDA scheme to enhance the discriminatory power of the obtained D-LDA feature space. The effectiveness of the proposed method has been demonstrated through experimentation using two popular face databases [3]. The DF-LDA method presented here is a linear pattern recognition method. Compared with nonlinear models, a linear model is rather robust against noises and most likely will not over fit. Although it has been shown that distribution of face patterns is highly non convex and complex in most cases, linear methods are still able to provide cost effective solutions to the FR tasks through integration with other strategies, such as the principle of "divide and conquer" in which a large and nonlinear problem is divided into a few smaller and local linear sub-problems [1].

The photo-to-sketch transformation method was effective approach for automatic matching between a photo and sketch. This reduced the difference between photos and sketches significantly and increased the effective matching between the two. The recognition performance was even better than that of human beings. Without considering the absolute recognition accuracy the relative superior performance of the new method compared to human performance and the conventional photo based methods made this method advantageous. In this system geometric measures and Eigen face method are used. The geometrical feature method is intuitively the most straightforward method. A great amount of geometrical face recognition researches focus on extracting relative positions and other parameters of face components such as eyes, mouth and chin [4-6]. Although the geometrical features are easy to understand they do not seem to contain enough information for stable face recognition. In particular, geometrical features change with different facial expressions and scales, thus vary greatly for different images of the same person. The Eigen face method is a classic face recognition method. The disadvantage of this method is its sensitive to illumination, expression and rotation changes [5].

A system to search image using sketches was introduced. This involves two-phase method namely; sketch-to-mug-shot matching and human face image searching using relevance feedback. In the sketch-to mug-shot matching phase, a facial feature matching algorithm using local and global features. A point distribution model is employed to represent local facial features while global features consist of a set of geometrical relationship between facial features. The performance of this matching is good if the sketch image looks like the mug-shot image in the database. However in some situations it is hard to construct a sketch that looks like the photograph. To overcome this limitation this system used human-in-the-loop concept and image searching algorithm is proposed using relevance feedback in the second phase. Positive or negative samples will be collected from the user. A feedback algorithm that employs subspace linear discriminant analysis for online learning of the optimal projection for face representation is then designed and developed. Although the proposed system prototype woks well there are still some limitations on the proposed system. First, the current system detects the boundary of the facial features in a semi automatic manner which is quite time consuming for a large database. An automatic boundary detection algorithm is required when it is deployed into real application. Second the size of the testing database is relatively small compared with that in the real application. It is required to further evaluate the system performance in a large database. Finally the proposed feedback learning algorithm is based on LDA. If the feedback samples are non-linearly distributed, the systems performance may be degraded [3].

A novel face photo sketch synthesis and recognition method using a multi-scale Markov random fields (MRF) model has been introduced. This system has three components: 1.Given a face photo, synthesizing a sketch drawing 2.Given a face sketch drawing, synthesizing a photo.3.Searching for face photos in the database based on a query sketch drawn by an artist. It has useful applications for both digital entertainment and law enforcement. The faces are studied in frontal pose with normal lighting and neutral expression and have no occlusions. To synthesize sketch/photo images, the face region is divided into overlapping patches for learning. The size of the patches decides the scale of local face structures to be learnt. From the training set which contains photo-sketch pairs, the joint photo-sketch model is learnt at multiple scales using a multi-scale MRF model. By transforming a face-photo to a sketch or transforming a sketch to a photo, the difference between photos and sketches is significantly reduced, thus allowing effective matching between the two in face-sketch recognition. After the photo-sketch transformation in principle, most of the proposed face-photo recognition approaches can be applied to face-sketch recognition in a straight forward way. Since we assume all the faces are to be in the frontal pose with normal lighting and neutral expression and have occlusions this algorithm is efficient. If the photos in the training set are different from the input photo, the sketch-synthesis algorithm may not work well with significant variations [4].

Algorithm: The algorithms used in this paper are shot boundary Detection, sub shot boundary detection and Principle Component Analysis (PCA).

Shot Boundary Detection: There are three most commonly used transitions in web videos. They are hard cut, fade and dissolve. The explanation of the three transitions is as follows.

Hard Cut: The hard cut is the most important transition, especially for unedited video material and describes the abrupt change between two shots. We use color histograms to detect hard cut. For noise robustness, we reduce the number of histogram bins in HSV color spaces to 256 bins according to MPEG-7 Scalable Color (SCD). An adaptive threshold tha(n) is calculated independently for each area:

$$th_a(n) = a\left[\left(\sum_{m=n-N}^{n+N-1} D_a(m, m-1)\right) - D_a(n, n-1)\right] + 1 \quad (1)$$

A hard cut is detected if

$D_a(m, m-1) > th_a(n)$ and $D_a(n, n-1) < th_a(n+1)$ is true for all areas (a=1…3).

Fade: The disappearing shot fades into a blank frame followed by the appearing shot that fades in. The luminance average Yμ and variance $Y\sigma^2$ exhibit a certain pattern which is insensitive to noise and camera motion. So we detect the centre of fades by thresholding the strictly monotonic decreasing first derivative of the luminance variance.

Dissolve: A dissolve is defined by a temporal overlap of a few frames of the disappearing and appearing shot. Dissolve candidates are extracted by the characteristics of the first and second derivatives of $Y\sigma^2$.b Due to low image quality and fast camera operations, this approach produces many dissolve candidates. Our efforts lie in the verification of these candidates.

Sub Shot Boundary Detection: The segmentation of shots in smaller units has a special importance for single shot videos and for long shots. This division into sub shots targets to detect new visual content appearing through camera operations or special effects. The candidates for subdividing are determined by a location memory model and the camera motion. The location memory model is based on accumulated color histograms H in the HSV color space. At each point in time, this model computes the normalized distance distloc(n)

$$dist_{loc}(n) = \sum_{k=0}^{k-1}\left|H_n(k) - \frac{1}{N-1}\sum\right. \quad (2)$$

Extraction of Key Frames: Our key frame extraction approach targets to get a representative frame of a (sub)shot with noticeable visual content and in best possible quality. The motion intensity fmi calculated over all M ×N motion vectors in compressed domain is proportional to blurring effects in each frame:

$$\sqrt{f\left(\frac{1}{M}\sum_{i=0}^{M-1} mv_x(i)\right)^2 + \left(\frac{1}{N}\sum_{i=0}^{N-1} m\right)}$$

Camera motion is an important tool for the director of a video. ”Zooming in” attracts attention to details while ”zooming out” emphasizes the surroundings. The camera motion fcm is defined as:

□ □

fcm is positive during zooming, negative during translational camera and zero for fixed camera operations. The appearance of face and text also increases the attraction of such frames.

A useful parameter is the area of faces F:

$$f_{face}(n) = \sum_{i=1}^{F} f_{face_{width}}(i) \times f_{face_h} \square$$

The parameter ftext is a binary value according to the text detection 2

$$f_{text}(n) = \begin{cases} 1, te \\ 0, o_i \end{cases} \tag{6}$$

In order to avoid inexpressive and visually similar key frames, we further integrate the normalised variance of luminance fYσ2 and the difference on Color Layout fdiff CL to the last extracted key frame fcl (n =nlk)

$$f_{dif\ f_{CL}}(n) = \sqrt{\sum_{K=0}^{K-1}\left[f_{cl(k)}\,|\,(n) - f_{CL}(\square)\right]}$$

A key frame is detected by using the following formula $f_{att}(n)$

$$f_{att}(n) = \omega_{mi} \cdot f_{mi}(n) + \omega_{cm} \cdot f_{cm}(n) + \omega_{text} \cdot f_{text}(n) + \omega_{face} \cdot f_{face}(n) + \omega\gamma_{\sigma^2} \cdot (n) + \omega_{diff_{CL}} \cdot f_{diff_{CL}}(n)$$

Principle Component Analysis: The Principle Component Analysis (PCA) algorithm is as follows

Step 1: Calculate the mean image.

Step 2: Calculate the deviation of each image from mean image.

Step 3: Construct the covariance matrix.

Step 4: Find the Eigen values and Eigen vectors of the covariance matrix.

Step 5: Project the Eigen values in face space for face recognition.

The Mathematical Expression of Pca as Follows

Step1: $\bar{x} = \frac{1}{M}\sum_{I=}^{M}$

Step 2: Subtract the mean $\varnothing_i = x_i$.

Step 3: Form the matrix A=[$\varnothing_1$, $\varnothing_2$ ...]. Compute,

$C = \frac{1}{M}\sum_{n=1}^{M}\varnothing_n\varnothing_n^T = \frac{1}{M}A$ (Covariance matrix)

Step 4: Compute the Eigen values of C: 1 >2> .....>

Step 5: Compute the Eigen vectors of C: μ_2,....

Step 6: (Dimensionality reduction step) Keep only the terms corresponding to the K largest Eigen values.

$$\hat{x} - \bar{x} = \sum_{i=1}^{k} b_i \mu_i$$

System Design: System design has 2 modules. They are key frame extraction, Implementing PCA algorithm.

Key Frame Extraction: Key frame is the frame which can represent the salient content and information of the shot. The key frames extracted must summarize the characteristics of the video and the image characteristics of a video can be tracked by all the key frames in time sequence. Furthermore, the content of the video can be recognized. A basic rule of key frame extraction is that key frame extraction would rather be wrong than not enough. So it is necessary to discard the frames with repetitive or redundant information during the extraction [5-10].

Video segmentation and key frame extraction are the bases of video analysis and content-based video retrieval. Key frame extraction is an essential part in video analysis and management, providing a suitable video summarization for video indexing, browsing and retrieval. The use of key frames reduces the amount of data required in video indexing and provides the framework for dealing with the video content [8-2].

For video, a common first step is to segment the videos into temporal "shots," each representing an event or continuous sequence of actions. A shot represents a sequence of frames captured from a unique and continuous record from a camera. Then key frames are to be extracted. Video segmentation is the premise of key frame extraction and key frames are the salient content of the video (key factors to describe the video contents).

A video consists of video sequences. The video sequences are segmented into shots and sub shots. The shots and sub shot have number of frames. The key frames are extracted from the frame sequence.

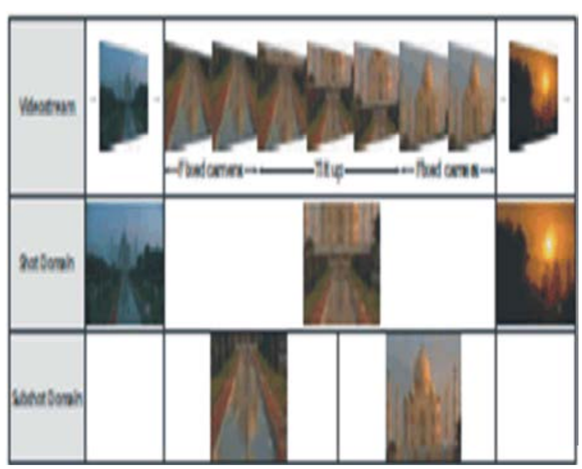

Fig 1: KeyFrameExtraction

Implementing PCA Algorithm: PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a "shadow" of this object when viewed from its (in some sense) most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced [6]. There are three modules in implementing PCA algorithm. They are Creation of training database, EigenFaceCore, Recognition.

Creation of Training Database: The faces in the extracted key frame are stored in the database. They are converted into gray scale images. The 2-D images are converted into 1-D vector images. Then the 1-D vector images are concatenated into a single matrix T.

Eigen Face Core: The output of first module is given as the input of second module. First the mean image is calculated. Then the deviation of the image is calculated. The centered images are merged into a single matrix A. By using A covariance matrix L is calculated L=A'*A. Eigen values and Eigen vectors are calculated. Eigen values are stored as diagonal elements. A threshold is set. The Eigen values greater than the threshold is considered. Eigen vectors is calculated by using the formula A*Eigen value. The query image is also considered. Eigen vectors are calculated for both the faces and query image. PCA algorithm is applied and the threshold values are calculated.

Face Recognition: The images are projected into face space. The projected image is calculated using the formula Eigen Faces' *A. The test image is considered. Then the test image is also projected in face space. Projected test image is calculated using the formula Eigen faces' *Diff Image(calculated from test image). Finally the Euclidean distances between the projected image and projected test image are calculated.

CONCLUSION

Facial recognition is most beneficial to use for facial authentication than for identification purposes, as it is too easy for someone trace, features with a disguise or mask, etc.This paper represents the face recognition in videos using the key frame extraction. Its mainly used for forensic department and criminal investigations. By using this method the faces can be recognized even from the video sequences.

REFERENCES

1. Juwei Lu, K.N. Plataniotis and A.N. Venetsanopoulos, 2002. Face Recognition using LDA based algorithm" IEEE transaction.
2. Pascal Kelm, Sebastian Schmiedeke and Thomas Sikora, 0000. Feature-Based Video Key Frame Extraction For Low Quality Sequences.
3. Pong C. Yuen and C.H. Man, 2007. Human Face Image Searching System Using Sketches" IEEE Transaction.
4. Xiaogang Wang and Xiaoou Tang, 2009. Face Photo-sketch synthesis and face Recognition "IEEE Transaction.
5. Xiaoou Tang and Xiaogang Wang, 2004. Face sketch Recognition "IEEE Transaction,
6. Yong Zhang, Christine McCullough, John R. Sullins and Christine R. Ross, 2010. Hand-Drawn Face Sketch Recognition by Humans and a PCA-based algorithm for Forensic applications".
7. Kishwar Sultana, Najm ul Hassan Khan and Khadija Shahid, 2013. Efficient Solvent Free Synthesis and X Ray Crystal Structure of Some Cyclic Moieties Containing N-Aryl Imide and Amide, Middle-East Journal of Scientific Research, 18(4): 438-443.
8. Pattanayak Monalisa and P.L. Nayak, 2013. Green Synthesis of Gold Nanoparticles Using Elettaria cardamomum (ELAICHI) Aqueous Extract World Journal of Nano Science & Technology, 2(1): 01-05.
9. Chahataray Rajashree and P.L. Nayak, 2013. Synthesis and Characterization of Conducting Polymers Multi Walled Carbon Nanotube-Chitosan Composites Coupled with Poly (P-Aminophenol) World Journal of Nano Science & Technology, 2(1): 18-25.
10. Parida U.K., S.K. Biswal, P.L. Nayak and B.K. Bindhani, 2013. Gold Nano Particles for Biomedical Applications, World Journal of Nano Science & Technology, 2(1): 47-57.