Journal of Information Systems and Communication
ISSN: 0976-8742 & E-ISSN: 0976-8750, Volume 3, Issue 1, 2012, pp.154-157.
Available online at http://www.bioinfo.in/contents.php?id=45

TOWARDS RETRIEVAL OF HUMAN FACE FROM VIDEO DATABASE: A NOVEL FRAMEWORK

ZAFAR G. SHEIKH[1], THAKARE V.M. [2] AND SHEREKAR S.S. [3]

Dept. of Computer Science and Engg., SGB Amravati University, Amravati, MS, India
*Corresponding author: E-mail- zgsheikh@gmail.com, vilthakare@yahoo.co.in, ss_sherekar@rediffmail.com

Received: January 12, 2012; Accepted: February 15, 2012

Abstract- With enormous growth of security and surveillance system, a huge amount of video data is being generated every day. It is immense challenge for researcher to search and retrieve human face of interest from video. The proposed work is inspired from the same issue. It would be the future demand for searching, browsing, and retrieving human face of interest from video database for several applications. This paper proposed the novel framework for human face retrieval from video database using frontal face image as a query to video database. Human face is detected using Viola and Jones frontal detector. The features were extracted using fast Haar features based algorithm. At the same time, the grouping of detected faces within one video sequence shot into face tracks using Kanade-Lucas-Tomasi (KLT) tracker. The selected features were match with face tracks for face recognition.
Keywords- face retrieval, face recognition, face detection, face tracking

Citation: Zafar G. Sheikh, Thakare V.M. and Sherekar S.S. (2012) Towards Retrieval of Human Face from Video Database: A Novel Framework. Journal of Information Systems and Communication, ISSN: 0976-8742 & E-ISSN: 0976-8750, Volume 3, Issue 1, pp-154-157.

Introduction

Human face detection and recognition from video database is very intuitive to computer and human [1], still various challenges to computer for face detection described in [2] like pose, scale, illumination, expression etc. Work is in progress to overcome the problems in real time applications. The objective of this paper is to build an framework for retrieve the human face from the video database using selected features, under the assumption that face is independent of the above problems i.e. explore the human face retrieval with (near) frontal face. Based on the proposed work, users can easily acquire the information that, interested human face image is available in video or not. If it is available, then retrieve the corresponding key frames from video. This framework is not only bring a new browsing and searching experience, but also provide an alternative of video summarization.

This paper presents a novel framework for human face retrieval from video. Videos have different categories like feature-length films, news videos, and surveillance and family videos. A human face image is work as query by detecting and extracting features. The selected features are matched with key frames. The multiple face detection from video is worked with scene and key frame extraction methods for efficient detection.

This framework involves four steps for human face retrieval, in first step uses Viola- Jones detector for face image detection. Second step for extracted and selected using fast Haar like features based algorithm, at the same time third steps to track face in video using Kanade-Lucas-Tomasi (KLT) tracker to track the interest point throughout the video frames, last and important step is face recognition using selected feature matched.

The remainder of this paper is organized as follows. Section II described the related work, the proposed framework appears in Section III, and Section IV concludes with future work.

Related work

Image retrieval based applications on visual content such as QBIC, Netra, VisualSeek, WebSeek, Virage, VideoQ, MARS are available for use. The limitations and challenges were discussed in [3] related with image/video searching and retrieving closely associated with CBIR. Google is also work on object matching in video [4] with text retrieval approach using SIFT descriptor for

view invariant. Soft biometrics is applied [5] for facial marks identification on FERET database using Active Appearance model (AAM) for improving face matching and retrieval. Whereas, Josef Sivic et al [6] efforts to find all occurrence of a particular person in shot with changes in scale, pose and partially occlude using Gaussian mixture modal in RGB colour space. New people detected in video stream [7] by Viola and Jones face detector and a kernel based regressor face tracking. Although Mark Everingham and A. Zisserman uses combination of generative and discriminative head models for identifying individuals in video [8],[9] 3-D ellipsoid approximation, [10] coarse 3-D model with multiple texture for character identification in situation comedies or feature-length films.

While the fusion of face and naming approaches were used for retrieval from videos. [11] Proposed a readily available texture source, the film script, which contain character name in front of their spoken lines. Yi-Fan Zhang et al. [12], applied global matching between names and clustered face tracks with association network. Towards person Google [13], is the combination of face detection and speaker segmentation for multimodal person retrieval using statistical normalization PCA. Similarly, [14] shows the framework for retrieve faces in the TV show video frame sequence.

The efforts which are more relevant to proposed approach such as, O. Arandjelovic and A. Zisserman [15] used face image as a query to retrieve particular characters. Affine warping and illumination correcting were utilized to alleviate the effect of pose and illumination variations. Whereas, [16] is proposed kernel-based SVM for visual feature retrieving actors in films. To overcome the problems in content based image retrieval, Pablo Navarrete et al. [17] is projected interactive face retrieval system using self–organizing maps. An integration of statistical and structural information [18] for the local feature constructed from coefficient of quantized block transforms. DCT features were used in [19] for face image retrieval based on centered position of two eyes. Chon Fong Wong et al [20] is employed Adaboost based face detection and Lifting Wavelet Transform (LFWT) for feature extraction for in video sequences. The [21] utilized intelligent fast-forwards to jump video to the next scene containing that face, affine covariant region tracker for face region tracking. Recently, geometrical face attributes using shape manipulation [32] is used for face image retrieval.

Proposed framework

The proposed framework (Figure 1) involves four steps for human face retrieval, in first step Viola-Jones detector use for face image detection. Scene detection is performed in video database using fast-forward method and key frame extraction with Latent Aspect Modeling (LAM). Second step for extracting and selecting features using fast Haar like features, then in third steps to track face in video using Kanade-Lucas-Tomasi (KLT) tracker to track the interest point throughout the video frames, last and important step normalization and likelihood for is face recognition

A. Face Detection

From given arbitrary image, the goal of face detection is to determine whether or not there are any faces in the image and if present return the image location and extent of each face. Problem is challenging because faces are nonrigid and have a high degree of variability in size, shape, color, and texture. Numerous techniques have been developed to detect faces in a single image, [2] had been provides detail survey of face image detection algorithms, categorize and evaluate these algorithms. Whereas, Ming-Hsuan Yang et al [23] had been covered early works (before 2004) of recent advancement in face detection. Recently (In 2010) Microsoft Research [22] released a technical report on the same issue.

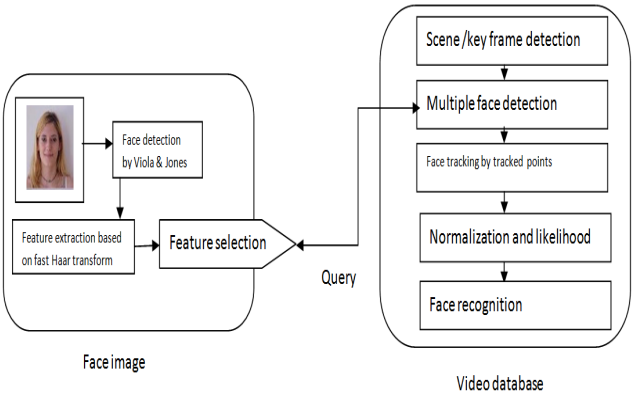

Fig. 1- Diagrammatic representation of proposed human face retrieval framework

The Viola-Jones Face Detector

Three main ideas that make it possible to build a successful face detector [25] that can run in real time: the integral image, classifier learning with AdaBoost, and the attentional cascade structure. Viola-Jones introduced [24] a new image representation called as "Integral Image" for rapid computation of Haar-like features, as detailed below.

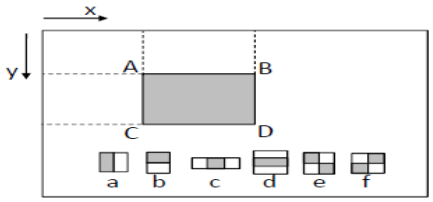

Fig. 2- The integral image and Haar-like rectangle features (a-f)

The integral image is constructed as follows:

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x', y'), \qquad (1)$$

Where $ii(x,y)$ the integral is image at pixel location (x,y) and $i(x',y')$ is the original image. Using the integral image to compute the sum of any rectangular area is extremely efficient, as shown in Fig. 2. The sum of pixels in rectangle region ABCD can be calculated as:

$$\sum_{(x,y)\in ABCD} i(x,y) = ii(D) + ii(A) - ii(B) - ii(C), \quad (2)$$

Which is only requires four array references.

The integral image can be used to compute simple Haar-like rectangular features, as shown in Figure (2) (a-f). The features are defined as the (weighted) intensity difference between two to four

rectangles. For instance, in feature (a), the feature value is the difference in average pixel value in the gray and white rectangles. Since the rectangles share corners, the computation of two rectangle features (a and b) requires six array references, the three rectangle features (c and d) requires eight array references, and the four rectangle features (e and f) requires nine array references.

Fig. 3- Viola-Jones Face detector's results

A simple and efficient classifier which is built using the AdaBoost learning algorithm to select a small number of critical visual features from a very large set of potential features. By combining classifiers in a "cascade" which allows background regions of the image to be quickly discarded while spending more computation on promising face-like regions.

Scene and Key Frame Extraction

The objective of this paper is to design such framework which is retrieving frames containing particular frontal face in video using a human frontal face image as the query. For minimizing the processing, proposed framework is performed scene detection. There are many applications [21] with capability called as intelligent fast-forward, the video jump to the next scene containing that particular person.

The automatic significant key frame selection and extraction is another issue. Many researchers have been working on extracting meaningful key frame for various applications. Jiebo Luo et al. [26] provided literature review on key frame extraction as well an automatic key frame extraction method. The proposed approach uses probabilistic Latent Aspect Modeling [27] over the possible local matches key frame image set. This allows the extraction of significant group of local matching descriptors that may represents characteristic elements of key-place.

B. Feature Extraction and Selection

For the face representation and recognition meaningful feature always play an important role. Features will affect the performance of retrieval and recognition of an object. Yanwei Pang et al. [31] have been published very effective literature survey for different available feature extraction algorithm such as Principal Component Analysis (PCA), Linear discriminant Analysis (LDA), locality preserving projection (LPP) , neighborhood preserving projection (NPP), marginal Fisher analysis (MFA) , orthogonal complement component analysis(OCCA) , geometric mean based learning , tensor projections, and spectral regression discriminant analysis (SRDA). To extract features with less computation cost authors proposed two effective algorithm for feature extraction, fast Haar transform based PCA (FHT-PCA) and fast Haar Transform based SRDA (FHT-SRDA).

The experimental results on ORL, Yale, and FERET database shows accelerate subspace analysis and feature extraction using two effective subspace based Fast Haar Transform algorithm. The proposed framework could conveniently adopt one of the effective feature extraction algorithm based on fast Haar Transform.

C. Face Tracking

From given set of faces detected with viola-Jones frontal face detector (fig. 4) [24], the problem is to group these faces into face tracks, so that, each track represents one unique person from video.

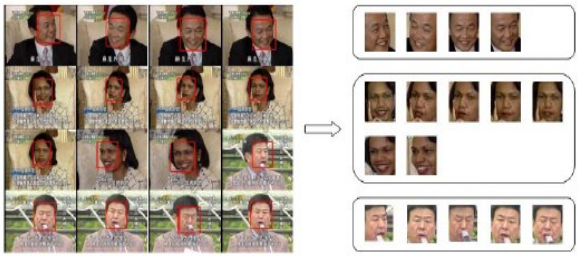

Fig.4- finding face track in input video sequence and detected face tracks

To group detected faces within one video sequence shot into face tracks consist of different facial expression of unique individual. Interest points can be detected from given pair of from images containing faces and tracked using [28] Kanade-Lucas-Tomasi (KLT) tracker. The interest point are selected according to textured criterion found in facial region under the assumption that there is not much difference in feature appearance and position, these points can be tracked using image motion model.

By counting the number of shared points and total number of points of pair of faces in each frames, a threshold used to decide a match.

Figure (5) shows demonstration grouping faces using Kanade-Lucas-Tomasi (KLT) tracker. However, face detector are inconsistent due to change in illumination, pose variation, facial expressions, and occlusions in actual video sequence.

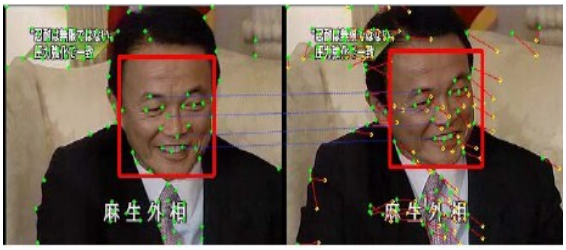

Fig. 5- Interest point using Kanade-Lucas-Tomasi(KLT) tracker

The tracker is tested on different news video broadcasting stations. The Viola–Jones face detector implemented in OpenCV was used for detecting frontal faces in every frame of video sequences. The total face tracks detected is 94.17%, which is better as compare to baseline method and Everingham et al method. Recently, [33] multi camera occlusion robust face tracking provides superior results.

Face Recognition

During the past decades face recognition in video has received significant attention. Current face recognition systems have reached a certain level of development, still pose, illumination, facial expression and low resolution problem unsolved. The scope of this paper is to retrieve frontal face exclusive of the above mentioned problems.

Face recognized by computer using approaches [1] like holistic templates, feature geometry, and hybrid for face identification and verification. Whereas, Huafeng Wang et al [29] were categorized face recognition in to spatial-temporal based, statistic model, and hybrid cues with detail survey of video based face recognition. It also covered the advanced topics like pose, illumination, and 3-D representation of face.

Viola-Jones is detecting frontal faces in video frame and KLT tracker has grouped the face tracks in video sequence. However, the features of detected face image were extracted and selected using Fast Haar transform based algorithm. The likelihood between the extracted features was matched with the normalized features from group of face track for face recognition. It can also performed by histogram equalization. [50] New distance measure from one feature from the one of the gallery image. Image-to-class distance is used from the set of local features for recognition of face and human gait.

Conclusion

It will be the future demand for searching, browsing and retrieving human face of interest from video database for several applications. The proposed paper provides a futuristic framework for human frontal face retrieval from video database using frontal face image as a query.

This framework is utilized Viola-Jones frontal face detector for detecting faces which is implemented in OpenCV. Interest point on face were track using KLT tracker from video sequence and features extracted and selected using Fast Haar transform based algorithm. These features were used for face recognition. The overall approach would be able to retrieve face frames from video on face image query, still implementation needs extra effort. The proposed framework is theoretical approach and therefore the cost of implementation will be calculated in the future after the actual implementation of framework.

References

[1] Rama Chellappa, Pawan Sinha and Jonathon Phillips P. (2010) *International Journals of IEEE Computer Society*.

[2] Ming-Hsuan Yang, David J. Kriegman and Narendra Ahuja (2002) IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(1), 36-58.

[3] Nida Aslam et al (2009) *International workshop on Multimedia information retrieval*.

[4] Josef Sivic and Andrew Zisserman (2006) *British Machine Vision Conference*.

[5] Unsang Park and Anil K. Jain (2010) *IEEE Transaction on Information forensics and security*.

[6] Josef Sivic et al. (2006) *British Machine Vision Conference*.

[7] Nicholas Apostoloff et al. (2007) *British Machine Vision Conference*.

[8] Mark Everingham et al. (2005) *International Conference on Computer Vision*.

[9] Mark Everingham et al. (2004) *British Machine Vision Conference*.

[10] Mark Everingham and Andrew Zisserman (2004) *International Conference on Image and Video Retrieval*.

[11] Mark Everingham et al. (2006) *British Machine Vision Conference*.

[12] Yi-Fan Zhang, Changsheng Xu et al. (2009) *IEEE transaction on multimedia*, 11(7).

[13] Lutz Goldmann, Amjad Samour and Thomas Sikora (2008) *Towards Person Google: Multimodal Person Search and Retrieval, Technical report*.

[14] Yina Han et al. (2008) 2*nd International conference the K-Space*.

[15] Ognjen Arandjelovie et al. (2005) *IEEE Conference on Computer Vision and Pattern Recognition*.

[16] Shuji Zhao et al (2008) 16*th European signal processing conference*.

[17] Navarrete P. and Ruiz-Del-Solar (2002) *International Joint Conference on Neural Networks*, 687-691.

[18] Daidi Zhong and Irek Defee (2008) *EURASIP Journal on Advances in Signal Processing*.

[19] Aamer S.S., Mohamed et al. (2006) *An efficient face image retrieval through DCT features*.

[20] Chon Fong Wong et al. (2005) *Ninth IEEE International Symposium*.

[21] Josef Sivic et al. (2006) *International Journal of computer vision*.

[22] Cha Zhang and Zhengyou Zhang (2010) *Technical report by Microsoft Research*.

[23] Ming Hsuan Yang (2004) *ICPR*.

[24] Viola P. and Jones M. (2001) CVPR, 1-13.

[25] Viola P. and Jones M. (2004) *International Journal of Computer Vision* 57(2), 137-154.

[26] Jiebo Luo, Christophe Papin, Kathleen Costello (2009) *IEEE transaction on Circuits and systems for video tech.*,19 (2).

[27] Maguelonne Heritier, Langis Gagnon and Samuel Foucher (2009) *IEEE Transaction on circuit and systems for video technology,* 19(6).

[28] Thanh Duc Ngo et al (2008) *Signal Image Technology and Internet Based Systems, IEEE International Conference*.

[29] Huafeng Wnag, Yunhong and Yuan Cao (2009) *World Academy of Science, engineering and Technology*, 60, 293-302.

[30] Yi Huang, Dong Xu and tat-Jet Cham (2010) *IEEE Transaction on circuit and systems for video technology*, 20(3).

[31] Yanwei Pang, Xuelong Li, Yuan Yuan, Dacheng Tao and Jing Pan (2009) *IEEE Transaction on information forensics and security,* 4(3).

[32] Brandon M. Smith, Shengqi Zhu, Li Zhang (2011) *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[33] Josh Harguess, Changbo Hu and Aggarwal J.K. (2011) The 3rd *International Workshop on Machine Learning for Vision-based Motion Analysis*.