Methodology for Human Face retrieval from video sequences based on holistic approach

Z. G. Sheikh
Dept. of Computer Science
SGB Amravati University
Amravati

V.M. Thakare
Dept. of Computer Science
SGB Amravati University
Amravati

S. S. Sherekar
Dept. of Computer Science
SGB Amravati University
Amravati

ABSTRACT
Huge amount of video data is being generated every day, with enormous growth of security and surveillance system. It is immensely challengeable for researcher to search and retrieve accurate human face of interest from video with utmost speed. The proposed work is stimulated from the same concern. It would be the future demand for searching, browsing, and retrieving human face of interest from video database for several applications.
This paper proposes the novel methodology for human face retrieval from video database based on holistic approach. The Viola and Jones frontal face detector detect the face region and it is converted into 3-D ellipsoid model as a query to video database. At the same time, all the faces from video are detected and converted into 3-D model. The recognition is performed by using chamfer distance for each frame sequence. 3-D ellipsoid model has an advantage over face angle and facial expression variation. The performance evolution discusses the advantages of part-based approach.

General Terms
Computer Vision, pattern recognition, video processing

Keywords
Face detection, face recognition, tracking, 3-D ellipsoid

1. 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 the methodology for retrieve the human face from the video database using holistic approach, under the assumption that face is independent of the above problems i.e. explore the human face retrieval with (near) frontal face with some variation in face pose and facial expressions. 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 with face model. This model 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 three steps for human face retrieval, in first step uses Viola- Jones detector for face image detection. Second step for conversion of face image into 3-d ellipsoid model, at the same time multiple face detection from each frame of video and converted into 3-D ellipsoid model. A third and last step is to recognize the query face image from video sequences using chamfer distance.

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

2. 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 working 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] takes 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.

While the fusion of face and naming approaches were used for retrieval from videos. [8] Proposed a readily available texture source, the film script, which contain character name in front of their spoken lines. Yi-Fan Zhang et al [9], applied global matching between names and clustered face tracks with association network. Towards person Google [10], is the combination of face detection and speaker segmentation for multimodal person retrieval using statistical normalization PCA. Similarly, [11] shows the framework for retrieve faces in the TV show video frame sequence. In[12], we presents a frame for face retrieval using KLT tracker.

Whereas, O. Arandjelovic and A. Zisserman [13] are uses 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, [14] 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.*[15] is projected interactive face retrieval system using self–organizing maps. An integration of statistical and structural information [16] for the local feature constructed from coefficient of quantized block transforms. DCT features were used in [17] for face image retrieval based on centered position of two eyes. Chon Fong Wong *et al.* [18] is employed Adaboost based face detection and Lifting Wavelet Transform (LFWT) for feature extraction for in video sequences. The [19] utilized intelligent fast-forwards to jump video to the next scene containing that face, affine covariant region tracker for face region tracking.

The efforts which are more significant to propose approach such as, Mark Everingham and A. Zisserman uses combination of generative and discriminative head models for identifying individuals in video [20],[21] 3-D ellipsoid approximation, [22] coarse 3-D model with multiple texture for character identification in situation comedies or feature-length films.

3. PROPOSED METHODOLOGY

This model involves three stages (figure 1) for human face retrieval, in first stage uses Viola- Jones detector for face image detection. Second stage for conversion of face image into 3-d ellipsoid model, at the same time multiple face detection from each frame of video and converted into 3-D ellipsoid model. Final stage is to recognize the query face image from video sequences using chamfer distance.

3.1 Holistic Approach

The holistic approach generally refers to methods that use the entire face image for face identification. Basically it includes methods like Principal Component Analysis (PCA) or Eigenfaces, Linear Discriminate Analysis (LDA) or Fisherface. We consider the recent work using 3-D modeling of face and head.

3.2 The Viola-Jones Face Detector

Three main ideas that make it possible to build a successful face detector [23] 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.

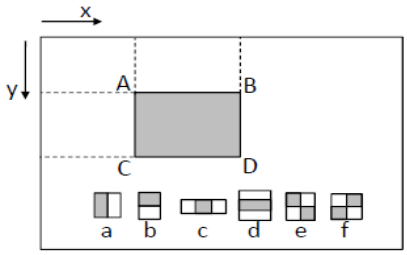

Figure 2- The integral image and Haar-like rectangle features (a-f)

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 figure 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.

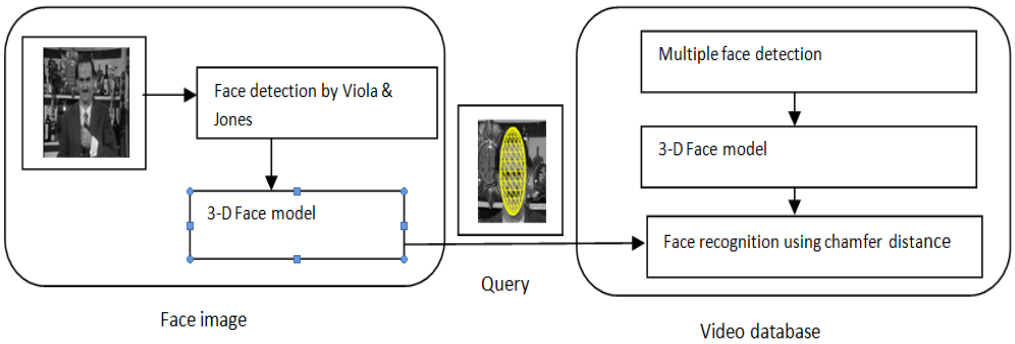

Fig. 1. Diagrammatic representation of proposed human face retrieval

Figure 3- Viola-Jones Face detector's results

Figure 3 shows face detectors results using Cascade classifier. 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.

3.3 3-D face Model

The pose based face rendering is performed by applying 3-D geometric model with multiple texture maps [20]. (Figure 4) shows training image (a) and face detected by well known detector (b) and the ellipsoid model generated(c). The single training image is back projected onto ellipsoid to given texture maps, can obtained new view of head move with different pose rendered by transforming the ellipsoid and projecting the texture maps back into the image. Multiple texture map can accurate rendering on many poses and differing appearance (facial expression). Pose is estimate and normalized with multiple appearances in-plane rotation with 6-D vector corresponding to rotation, scale and 2-D translation. Distance between poses is computed by the dot product between a front-facing vector normal to the ellipsoid. The representation of face image suitable for person classification performed using edge-based descriptor. Model is learned with probability distribution for the finding specific character in shot with variant in expression and pose.

The pose is parameterized as a 6-D vector p = (θ, φ, ψ, σ, τx, τy) corresponding to rotation, scale, and 2-D translation in the image. Rotation is specified by azimuth θ, elevation φ, and in-plane rotation ψ. This parameterization allows reasonable bounds to be specified easily.

A candidate face region provides an initial estimate of scale ˜σ, up to the scale step between pyramid levels, and translation (˜τx, ˜τy)(the centre of the candidate region). The task is to find the pose parameters ˆp which maximize the similarity between the rendered view R(p, μ) and the target image I. Normalized cross correlation(NCC), masked by the silhouette of the rendered view, is used as the similarity measure:

$$\hat{p} = \arg\max_{p}[\max_{\{\mu\in\mu_p\}} NCC(I, R(p,\mu))] \qquad (3)$$

For a given pose, multiple appearances R(p, μ) are proposed by selecting a subset of the texture maps {μp} which are (i) close to the current pose, and (ii) varying in expression. This is done by first finding the texture map which has pose q closest to the current estimate p, then selecting all texture maps with pose close to q (which represent different facial expressions)[20].

Whereas, the [21] presented an approach to locate the individual character frames of person with large changes in scale without using the temporal information. The approach work in two stages first is to data collected automatically by frontal face detection and clustering it given ellipsoid model for the character and the corresponding texture map. For the variation in shape and other part of appearance invariant translation model used. Then apply constellation model over each aspect of local maxima and search using image pyramid for the verification with likelihood of model. While training stage, choosing patches around interesting points in the texture maps. PCA based model by Gaussian with diagonal covariance use for variation the data. The second stage is for verification generated by constellation model by assuming affine camera and four position of the corresponding ellipsoid to determine the pose. A gray scale image of ellipsoid model is rendered in the estimated pose for comparison against the input image.

3.4 Face recognition using Chamfer Distance

For the pose refinement on obtained 3-D coordinates, the matching error to edge in the input image is defined as a robust directed chamfer distance and refined minimizing chamfer distance using LM-ICP algorithm which transform distance to make computing nearest edge efficient and Levenberg- Marquardt optimization. The final stage of recognition of particular character is defined with chamfer distance after pose refinement [22].

Fig 4. a) Training image b) face detected c) ellipsoid model

The matching error to edges in the input image is defined as a robust directed chamfer distance defined as-

$$d(U,V) = \frac{1}{|U|}\sum_{u_i\in U}\min\left(\min_{v_j\in V}\|u_i - v_j\|, \tau\right) \qquad (4)$$

Where U is the set of model edge points and V is the set of input image edge points. The confidence that detection is due to a particular character i is defined as-

$$C(i) = \left[\frac{d(U_i,V)}{\min\left(\min_{j\neq i} d(U_j,V), k\right)}\right]^{-1} \qquad (5)$$

Where $d(Ui, V)$ is the chamfer distance (4) Using the ratio between the distance to the character of interest and the nearest of the other characters gives a more informative score than the distance alone. The constant κ is introduced to reduce

false positives on characters other than those modeled, and non-face detections [22].
There are the alternatives for the chamfer distance for robust face recognition using Maximum Correntropy Criterion [25] and Isogeodesic Stripes [26] for 3-D face recognition.

4. RESULT ANALYSIS AND DISCUSSION

The proposed model based on holistic approached with face detector using cascaded AdaBoost. In [20],the task is to detect the frames contains each character, and identify the image position and pose of the face correctly. Pose of the ground truth faces in the video covers poses of around +/-60 azimuth, +/_30 elevation and +/_45 in-plane rotation. Faces vary in scale from 15 to 200 pixels. By using face detector in place of color segment would increases the rate of accuracy and speed.

Algorithm [21] tested on 1,500 key-frames taken one per second from the episode 'A Touch of class' of the BBC sitcom 'Fawlty Towers', detecting three main characters. The pose variation exceeding +/-30 about 3-D axis and correctly identify the character. The correct identification requires both detection and recognition of character, in contract to face detection or recognition. The Viola and Jones face detector [24] has been tested with 93% detection precision in video sequences. However, by applying the detector it would be achieve more proficiency than existing work. Whereas, [22] provides recall level of 50% the precision is around 80% for all characters in video sequence and by applying detector the speed of detection increases.

The part-based approach [12] which has been undergoes many stages and not considering the possibility of detection under the variation of pose and facial expression. The presented methodology works for invariant pose and facial expression with more precision rate.

5. CONCLUSION

The proposed methodology works for holistic approach with the objective of human face retrieval from video sequences. The whole face is the input to the video sequences as 3-D ellipsoid object. The 3-D face model has been generated with detected training image using back projection and rendering with texture maps. The chamfer distance is match in every frame sequence of the video. The more accurate and prominent face detector and recognition approaches increases the rate of precision and speed of retrieval.

6. REFERENCES

[1.] Rama Chellappa, Pawan Sinha, and P. Jonathon Phillips, *Face recognition by computers and humans*, International Journals of IEEE Computer Society, 2010

[2.] Ming-Hsuan Yang, David J. Kriegman and Narendra Ahuja, *Detecting Faces in Images: A Survey*, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 1, pp-36-58, JANUARY 2002

[3.] Nida Aslam et al, *Limitation and challenges: image/video search & retrieval*, International workshop on Multimedia information retrieval , Netherlands , 2009

[4.] Josef Sivic and Andrew Zisserman, *Video google: A Text Approach to object Matching in videos*, Proceedings of the British Machine Vision Conference, 2006

[5.] Unsang Park and Anil K. Jain , *Face matching and retrieval using soft biometrics*, International journal IEEE Transaction on Information forensics and security(TIFS) , 2010

[6.] Josef Sivic et al., *Finding people in repeated shots of the same scene*, proceeding of the British Machine Vision Conference, 2006

[7.] Nicholas Apostoloff et al., *Who are you? – Real time person identification*, Proceedings of the British Machine Vision Conference, 2007

[8.] Mark Everingham et al., *Hello! My name is… Buffy" – Automatic Naming of character in TV video*, Proceedings of the British Machine Vision Conference, 2006

[9.] Yi-Fan Zhang, Changsheng Xu et al, *Character Identification in feature length films using global face – name matching*, IEEE transaction on multimedia, Vol. 11 No. 7, Nov 2009

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

[11.] Yina Han et al, Speaker retrieval for TV show video by associating audio speaker recognition result to visual faces, Proceedings of 2nd International conference the K-Space, 2008

[12.] Zafar G. Sheikh, Dr. V. M. Thakare and Dr. S. S. Sherekar, "Towards Retrieval of Human Face from Video Database: A novel framework", International Journal of Information Systems and Communication, ISSN: 0976-8742 , E-ISSN: 0976-8750, Vol. 3, Issue 1, 2012

[13.] Ognjen Arandjelovie et al, Automatic *face recognition for film character retrieval in feature-length films*, In proceeding of IEEE Conference on Computer Vision and Pattern Recognition, San Diego , 2005

[14.] Shuji Zhao et al, Actor *retrieval system based on kernels on bags of bags*, 16th European signal processing conference, (EURASIP), Switzerland , Aug 25-29, 2008

[15.] Navarrete, P. and Ruiz-Del-Solar, *Interactive face retrieval using self-organizing maps* , Proceedings of the 2002 International Joint Conference on Neural Networks , IJCNN '02 , 687 - 691, 2002

[16.] Daidi Zhong and Irek Defee , *Face retrieval based on robust local features and statistical structure learning approach* , EURASIP Journal on Advances in Signal Processing , Volume 2008, Article ID 631297, 12 pages

[17.] Aamer S.S. Mohamed et al, *An efficient face image retrieval through DCT features*, In proc. of international conference, 2006

[18.] Chon Fong Wong et al, *Face image retrieval in video sequence using lifting wavelets transform feature extraction*, Proceedings of the Ninth IEEE International Symposium, 2005

[19.] Josef Sivic et al, Person *spotting: video shot retrieval for face sets*, International Journal of computer vision 2006 Springer Science + Business Media, 2006

[20.] Mark Everingham et al., *Identifying individuals in video by combining generative and discriminative head models*, Proceedings of the International Conference on Computer Vision, 2005

[21.] Mark Everingham et al., Automated *visual identification of character in situation comedies*, Proceedings of the British Machine Vision Conference, 2004

[22.] Mark Everingham and Andrew Zisserman, Automated *person identification in video*, Proceedings of the International Conference on Image and Video Retrieval, 2004

[23.] P. Viola and M. Jones, Robust Real-Time Face Detection, International Journal of Computer Vision 57(2), pp-137–154, 2004

[24.] P. Viola and M. Jones, Rapid object detection using boosted cascade of simple features, In Proc. of CVPR, pp-1-13, 2001

[25.] Ran He, Wei-Shi Zhengand Bao-Gang Hu, "Maximum Correntropy Criterion for Robust Face Recognition", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 33, NO. 8,PP. 1561-1576, AUGUST 2011

[26.] Stefano Berretti, Alberto Del Bimbo, and Pietro Pala, "3D Face Recognition Using Isogeodesic Stripes", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 12, PP. 2162-2177, DECEMBER 2010