ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 20, March 2017

Person retrieval from a synoptic video

C.Nagavani[1],M.Kanimozhi [2],K.S.Abinaya[3],N.Nivetha Muthulakshmi [4] ,A.Anto Amala [5]
Department Of Electronics and Communication Engineering,
Kamaraj College Of Engineering And Technology,Virudhunagar-01

***Abstract*—Person retrieval is a task, to accomplish tracking of a person in a video .Particular person has to be identified in a large volume of video data . To reduce the enormous volume of video data by using an object-based video synopsis and to extract the specific person very quickly through offline mode. The proposed methodology has three stages used ,they are key frame extraction , face detection and face recognition. Where key frame extraction is based on X^2 histogram, Face detection is done based on Viola Jones algorithm and Face recognition is done based on Facial feature descriptors.**

Keywords— Keyframe Extraction,Facial feature descriptors

I. INTRODUCTION

Detecting specific person from a video is a task, where whole video is need to be processed and retrieving the person from the large sum of data . Person retrieval from a synoptic video includes various steps they are, Key frame extraction, face detection and face recognition .Key-frame refers to the image frame in the video sequence which is representative and able to reflect the summary of a video content. By using the key-frame it is able to express the main content of video data clearly and reduce the amount of memory needed for video data processing and complexity greatly. So we could make the storage organization, retrieval and recognition of video information more convenient and efficient.

In this a method of computation of key-frame is done using thresholding of absolute difference of histogram is used. It computes thresholding point using mean and standard deviation of absolute difference of histogram difference of the consecutive frame. And Face detection also refers to the psychological process by which humans locate and attend to faces in a visual scene. Face-detection algorithms focus on the detection of frontal human faces.The Viola–Jones object detection framework can be trained to detect a variety of object classes, but it was motivated primarily by the problem of face detection.. Face recognition compare selected facial features from the image and a face database Some face recognition algorithms identify facial features by extracting landmarks, or features, such as relative position, size, and or shape of the eyes, nose, cheekbones, and jaw from an image of the subject's face.

The organization of paper is as following ,Section I gives the brief idea of proposed methodology .Section II Describes in detail the algorithm used in proposed system .Section III Results of proposed system are presented and discussed .Section IV Discuss the conclusion and issue related to proposed system .

II. PROPOSED METHODOLOGY

First, a lengthy video has to be processed. It is taken from a stored database. In that, a long duration video is converted into a short duration video. That short duration video is said to be a synoptic video. Synoptic video is obtained by extracting the keyframes from the long duration video. This synoptic video helps to recognize the face of a person quickly. After video synopsis, faces of the person are to be detected by Viola-Jones algorithm[13].Detected faces are cropped from the video and stored to a database .In face detection stage, a skin color based method is employed. It is to locate eye ,mouth and face boundary. Subsequently ,LBP and HOG features are extracted for this region.HOG and LBP features are combined to recognize the specific face. After recognizing the specific face, comparison is made. Comparison is done between the stored database and the query image .The sequence of frame containing the labelled faces displayed. Therefore ,informative video of a specific person can be taken for further analysis.

A. Keyframe Extraction

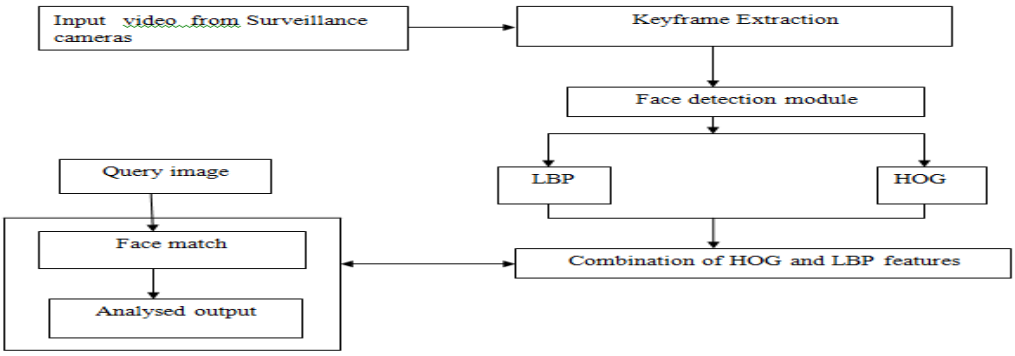
A video consists of many number of frames which will occupy large space and memory .Hence ,processing such long duration videos becomes a tedious process. Video is rich in content which results in a tremendous amount of data to be processed. This can be made easy by only processing few frames. Those few frames are keyframes which will reveal the content of the large video.

Keyframe extraction algorithm[9] is based on absolute differences of histogram of consecutive image frame. There are two phases in keyframe extraction. First

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 20, March 2017

,Compute the threshold using mean and standard deviation of histogram of absolute differences of consecutive image frames .Second phase is extracting keyframes one-by-one by comparing the threshold against absolute differences of consecutive image frames.



Keyframe extraction algorithm[9] is as follows:
Step 1: Extract frames one by one.
Step 2: Compute the X^2 histogram differences between two consecutive frames.
Step 3: Calculate the Mean and Standard deviation of absolute difference.
Step 4: Compute threshold.
Step5:Compare the difference with threshold and if it is greater than threshold, then it is a key frame else go to Step 2.
Step 6: Continue till end of the video.

B. Face Detection

The next stage in succession is face detection module which is used to detect the faces from the keyframe extracted .Viola and Jones method intended for real time application runs at 15frames/sec. There are four stages in Viola Jones algorithm. They are Haar feature selection, Creating an integral image, Adaboost training and Cascading classifiers.

Fig 1:Face detected from the video

Viola and Jones focus on detecting faces in images, but the framework can be used to train detectors for arbitrary objects, such as cars ,building. We can slide the fixed size window across our image at multiple scales .At each of these phases, our window stops ,computes some features ,and then classifies the region as yes, this region does not contain a face, or No, this region does not contain a face .For each of the stops along the sliding window path, five rectangular features are there. To obtain features for each of these five rectangular areas ,we subtract the sum of pixels from the white region from the sum of pixels from the black region. These five rectangular regions and their corresponding difference of sums ,we are able to form

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 20, March 2017

features that can classify parts of a face .Then ,for an entire dataset of features ,we use the Adaboost algorithm to select which ones correspond to facial regions of an image .To have compatibility, Viola Jones introduced the concept of cascades or stages. At each stop along the sliding window path ,the window must pass a series of tests where each subsequent test is more computationally expensive than the previous one .If any one test fails, the window is automatically discarded.Fig 1 shows the detected face from the video. [7] proposed a system in which the cross-diamond search algorithm employs two diamond search patterns (a large and small) and a halfway-stop technique. It finds small motion vectors with fewer search points than the DS algorithm while maintaining similar or even better search quality.

C.Histogram of Gradient(HOG)

HOG descriptors are mainly used in computer vision and machine learning for object detection. However, we can use HOG descriptors for quantifying and representing shape and texture.HOG has five stages namely,

1. Normalizing the image prior to description.
2. Computing gradients in both x and y directions.
3. Obtaining weighted votes in spatial and orientation cells.
4. Contrast normalizing overlapping spatial cells.
5. Collecting all Histograms of Oriented Gradients to form the final feature vector.

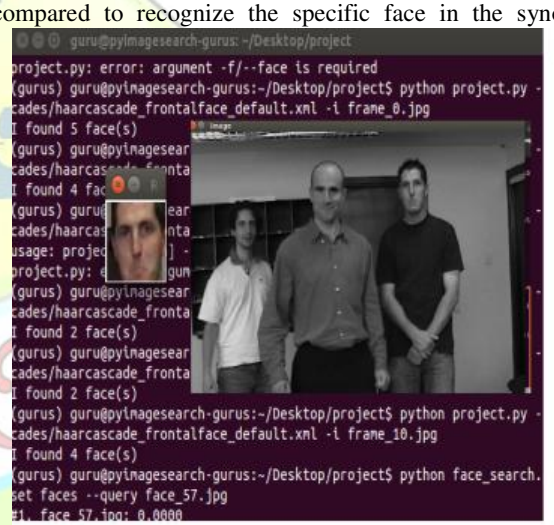
The reason HOG is utilized so heavily is because local object appearance and shape can be characterized using the distribution of local intensity gradients .However ,HOG captures local intensity gradients and edge directions ,it also makes for a good texture descriptor. The HOG returns a real valued feature vector. The cornerstone of the HOG descriptor algorithm is that the appearance of an object can be modelled by the distribution of intensity gradients inside rectangular regions of an image. Implementing the descriptor requires dividing the image into small connected regions called cells ,and then for each cell, computing the histogram of oriented gradients for the pixel within each cell. We can accumulate these histograms across multiple cells to form our feature vector.

D.Local Binary Pattern

Local Binary Patterns are used to characterize the texture and pattern of an image/object in an image. LBPs compute a local representation of texture. This local representation is performed by comparing each pixel with its surrounding neighbourhood of pixel values. The first step in constructing a LBP texture descriptor is to convert the image to gray scale. For each pixel in the gray scale image ,we select a neighbourhood of size r surrounding the center pixel .A LBP value is then calculated for this center pixel and stored in an output 2D array with the same width and height as the input image.

E.Face Detection :Combining HOG and LBP

SVM algorithm is used to separate strong features.HOG and LBP is combined and given to a SVM machine. The training samples are divided into two. The partition is done based on the threshold value. Feature with smallest classification error rate is chosen as root node. Again,it is spitted using another threshold. Sub-samples are stored and then classification is done based on SVM algorithm .For testing, feature vector of query image is compared to recognize the specific face in the synoptic



video. Query image is compared with the database face image so that the difference between the feature vector is measured .After recognizing ,it is compared with pre-stored database image. The resultant synoptic video contains only the specific person.

IV.RESULTS AND DISCUSIIONS

The experiment was carried with and without applying object based video synopsis. Table I shows the result of algorithm applied on videos available from various internet sources .The efficiency of the algorithm proposed has been evaluated using opencv3+python2.7. Video processes about 30frames/second for color images at different size.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 20, March 2017

The analysis is done using two ways,

1. Person Retrieving from a Non- synoptic video.
2. Person Retrieving from a Synoptic Video.

The direct parameters from the proposed system are for the analysis are the processing time (Pt), video time (Vt), frames per second (fps) and total number of frames. The derived parameters obtained from the direct parameters are the number of wrongly predicted frames and efficiency.

The number of wrongly predicted frames(WP) is calculated as given in Equation .1 the time of incorrect classification time
(IP_t), to the total time and total number of frames (F) of the video.

$$WP= IP_t*F / V_t \quad (1)$$

Similarly, the number of correctly predicted frames (CP) is calculated as given in Equation.2, by computing the time of correct classification time (CP_t), to the total time and total number of frames (F) of the video .

$$CP= CP_t*F /V_t \quad (2)$$

The efficiency of the system is computed by the following Equation.3

$$Efficiency= CP *100 /CP +WP \quad (3)$$

Table 1:Performance analysis of synoptic and Non-Synoptic Video

Process	Duration (minutes)	Wrongly predicted Rate (WP)	Correct Predicted Rate (CP)	Efficiency
Non-Synoptic Video	4.16	2	173	85%
Synoptic Video	2.06	1	174	92%

Fig 2: Person Retrieval from the synoptic video

*V.*CONCLUSION

Video synopsis has been proposed as an approach for condensing the activity in a video into a very short period.In order to recognize a specific person in a video,Viola Jones algorithm and HOG-LBP features are adopted.The goal is to observe much information in a short time,video synopsis serves this purpose.Specific person is detected by Viola Jones algorithm.HOG-LBP features are extracted for detected face.An SVM classifier is used to classify the strong and weak
features which helps in face recognition system.Finally,a specific person is recognized in a synoptic video even though if there is change in the face expression.In a non synoptic video,it takes much time to detect a specific person. using Viola Jones face detection algorithm and HOG-LBP face recognition algorithm.The proposed method shows the very high accuracy of 92% on face recognition system.

REFERENCES

[1] Pritch, Y., Rav-Acha, A., Peleg, S.: 'Non-chronological video synopsis and indexing', IEEE Trans. Pattern Anal. Mach. Intell., 2008, 30, (11), pp. 1971– 1984

[2] Zhu, Z., Gao, J., Yu, H.: 'Real-time face detection algorithm using fractal features in MPEG-4 video stream', J. Convergence Inf. Technol., 2012, 7, (6),pp .54-62

[3] Viola, P., Jones, M.: 'Robust real-time object detection', Int. J. Comput. Vis., 2004, 57, (2), pp. 137–154

[4] Nikolaidis, A., Pitas, I.: 'Facial feature extraction and pose determination', Pattern Recognit., 2000, 33, (11), pp. 1783–1791

[5] Huang, D., Shan, C., Wang, Y.: 'Local binary patterns and its application to facial image analysis: a survey', IEEE Trans. Syst. Man Cybern., 2011, 41, (6), pp. 765–781

[6] Rav-Acha, A., Pritch, Y., Peleg, S.: 'Making a long video short: dynamic video synopsis'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, New York, June 2006, pp. 435–441

[7] Christo Ananth, A.Sujitha Nandhini, A.Subha Shree, S.V.Ramyaa, J.Princess, "Fobe Algorithm for Video Processing", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), Vol. 3, Issue 3,March 2014 , pp 7569-7574

[8] Fischler, M., Elschlager, R.: 'The representation and matching of pictorial structures', IEEE Trans. Comput., 1973, 22, (1), pp. 67–92

[9] Sheena CV,N.K.Narayanan Keyframe extraction by analysis of histogram of video frame using Statiscal method,science direct 2015,pp. 41-50