Facial Retrieval System from Video

[1]Mahmoud Elgamal [2]Nasser Al-Biqami

1The Custodian of the Two Holy
Mosque Institute for Hajj Research,
Umm Al Qura University, Saudi Arabia

2The Custodian of the Two Holy
Mosque Institute for Hajj Research,
Umm Al Qura University, Saudi Arabia

ABSTRACT— ***A challenge problem for researchers to search and retrieve accurate human face in a large video database; which has a diverse of applications in security, surveillance systems,...etc. The paper proposes a search and retrieval system based on Viola-Jones face detection, this features extracted using fast Haar features algorithm, afterwards Kanad-Lucas-Tomasi(KLT)-tracker used to track and group the similar faces. The technique implemented using Matlab[9] and demo is shown.***

Keywords— Face detection and tracking, KLT-tracker.

1. INTRODUCTION

Human faces play an important role in efficiently indexing and accessing video contents, especially in large scale broadcasting news video databases. It is due to faces are associated to people who are related to key events and key activities happening from all over the world. There are many applications using face information as the key ingredient, for example, video mining, video indexing and retrieval, person identification and so on. However, face appearance in real environments exhibits many variations such as pose changes, facial expressions, aging, illumination changes, low resolution and occlusion, making it difficult for current state of the art face processing techniques to obtain reasonable retrieval results. Video surveillance has been evolving significantly over the years and is becoming a vital tool for many organizations for safety and security applications. Real-time Human face detection and recognition is very important and has many application areas, like security, airports, ...etc. Still various challenges to computer for face detection like pose, scale, illumination, expression etc[4, 8]. The purpose of this paper is to build an framework for tracking and detecting the human face from video using selected features, under the assumption that face is independent of the above problems i.e. detect the human face retrieval with (near) frontal face. Finally implementation of the technique is demonstrated. The paper is organized as follows: section(2) details of system architecture; Viola-Jones face detection, key frame extraction, feature extraction, and working example . Finally section(3) experiment and implemented demo.

2. SYSTEM ARCHITECTURE

The overall system architecture of the proposed algorithm for retrieving human-face is shown in figure (2.1); the system involves four stages for face retrieval. In the first stage Viola-Jones used to detect faces, scene detection is performed in video database using fast-forward method and key frame extraction with latent aspect modeling. In the second stage, Haar-like features used to extract and select features. In the third stage Kanade-Lukas-Tomasi(KLT) tracker used to track the interest point all over the video frames. In the fourth stage is the normalization and likelihood for face recognition.

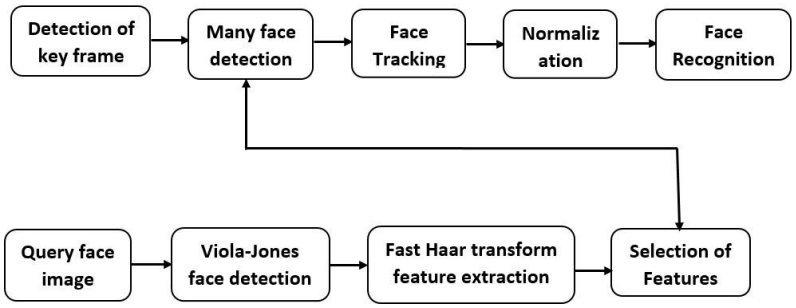


Figure 2.1: Face retrieval system.

2.1 Viola-Jones Face Detection

The Viola-Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time [7]. Similar to other previous methods, they used machine-learning algorithms to select a set of simple features which they combined into an efficient scalable classifier. The method introduced three key innovations that enabled their detector to achieve performance boosts over previous systems. The first was the use of the *integral image* for faster feature computation. The second was the use of the *AdaBoost* (adaptive boosting) machine learning algorithm as a means for quickly selecting simple and efficient classifiers. The third was a method for combining classifiers into a *cascade* to quickly eliminate background regions and focus computational attention on more promising areas of the image.

2.2 Haar-Like Features

The Viola-Jones face detection method uses combinations of simple Haar-Like features to classify faces. Haar-like features are rectangular digital image features that get their name from their similarity to Haar-wavelets; figure(2.2). Simple Haar-like features are composed of two adjacent rectangles, located at any scale and position within an image, and are referred to as *2-rectangle* features. The feature is defined as the difference between the sums of image intensities within each rectangle. Figure(2.1) illustrates the four different types of features used in the framework. The value of any given feature is always simply the sum of the pixels within clear rectangles subtracted from the sum of the pixels within shaded rectangles. Although they are sensitive to vertical and horizontal features, their feedback is considerably coarser. However, with the use of an image representation called the integral image, rectangular features can be evaluated in constant time, which gives them a considerable speed advantage over their more sophisticated relatives[2].

2.2.1 Integral Image

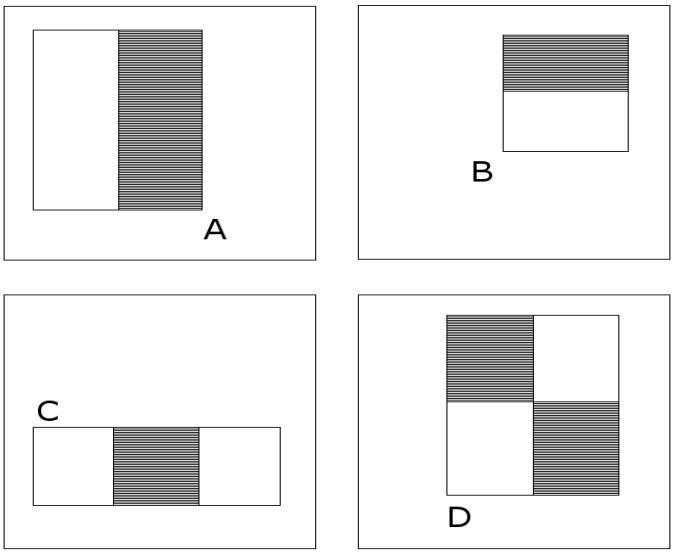
Haar-like features can be calculated extremely quickly by using an image representation called the integral image. The integral image is an application of summed-area tables, usually used in computer graphics. The integral image can be dened as:

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

Figure 2.2: Feature types used by Viola and Jones.

Where $ii(x,y)$ is the integral image and $i(x',y')$ is the original image intensity. The integral image can be calculated in a single pass using the following recurrences[2]:

$$s(x,y) = s(x,y-1) + i(x,y)$$
$$ii(x,y) = ii(x-1,y) + s(x,y)$$



Here $s(x,y)$ is the cumulative row sum and we have the following base cases:
$s(x,1) = 0$ and $ii(1,y) = 0$. Using the integral image, each feature can be calculated in constant time.

2.2.2 Cascade Architecture

The evaluation of the strong classifiers generated by the learning process can be done quickly, and arranged in a cascade in *order* of complexity. If at any stage in the cascade a classifier rejects the sub-window under inspection, no further processing is performed and continue on searching the next sub-window (see figure(2.3)). The effect of this single classifier is to reduce by roughly half the number of times the entire cascade is evaluated. The cascade architecture has interesting implications for the performance of the individual classifiers. Because the activation of each classifier depends entirely on the behavior of its predecessor, the false positive rate

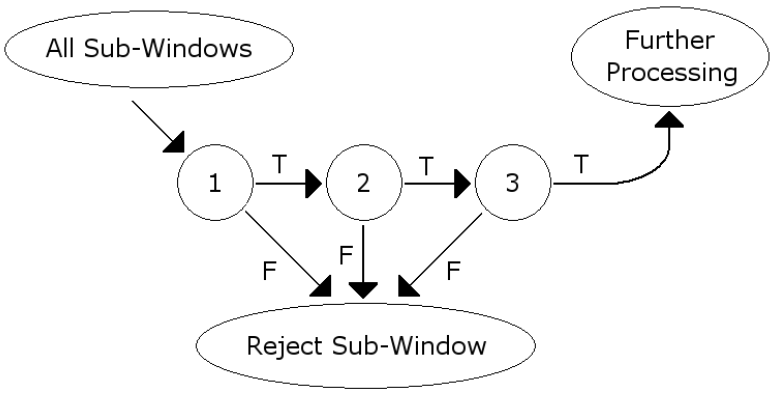


Figure 2.3: Cascade Architecture.

for an entire cascade is:

$$F = \prod_{i=1}^{k} f_i$$

Similarly, the detection rate is:

$$D = \prod_{i=1}^{k} d_i$$

Example for face detection using Viola-Jones algorithm is shown in figure (2.4) below.

2.3 Key Frame Extraction

The basic step for video analysis, efficient video indexing and content-based video retrieval is segmenting videos into scenes. 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. In this framework frame entropy and SURF descriptor used to find shot boundaries from the videos. Key frames were extracted from each shot and segmented the video into semantic scenes by key frame matching[1, 5], the flow diagram is shown in figure(2.7).

2.4 Feature Extraction

Significant feature plays crucial role for face representation and recognition.To extract features with less computational cost, two effective feature extraction algoFigure 2.4: Face detection using Viola-Jones.

Algorithms, termed fast Haar transform based PCA (FHT-PCA) and fast Haar transform based SRDA (FHT-SRDA). These two algorithms have two advantages: The computational costs for both the feature extraction process to obtain and the training process to approximate are very low with negligible or slight degradation of reconstruction error and recognition accuracy. These advantages make FHT-PCA and FHT-SRDA applicable to many practical subspace based applications[6].

Figure(2.6) shows algorithm for FHT-SRDA[6]. First, all the discriminant vectors are computed by SRDA. Then each discriminant vector is independently approximated by the Haar functions. Once the Haar functions and the corresponding coefficients are known, the low-dimensional features of the high-dimensional samples can be efciently extracted.

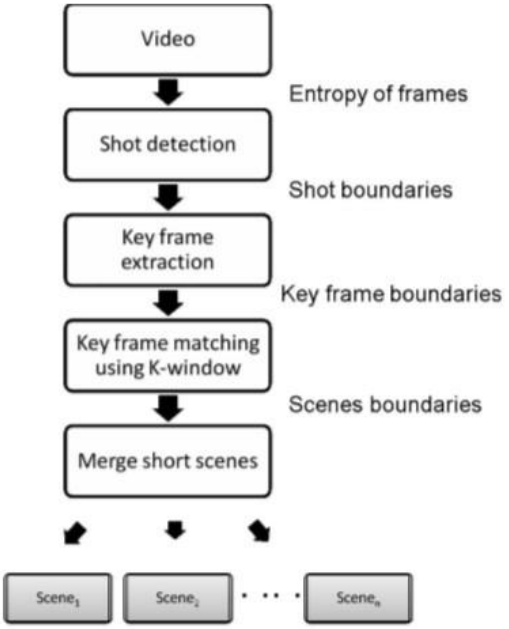


Figure 2.5: Segmenting Video into Scenes[1].

2.5 Detection and Tracking of Faces

Detecting and recognizing faces appearing in video are essential tasks of many video indexing and retrieval applications. After applying Viola-Jones face detection technique, it remain to group these faces into tracks which is a challenging problem. In order to make the face retrieval robust, the faces of the same person appearing in individual shots are grouped into a single face track by using a reliable tracking method. The retrieval is done by computing the similarity between face tracks in the databases and the input face track. For each face track, one representative face was selected and the similarity between two face tracks is the similarity between their two representative faces. The representative face is the mean face of a subset selected from the original face track using KLT-tracker[3]. KLT-tracker algorithm is shown in figure(2.7). The tracker tested on various videos, the Viola-Jones face detector used to detect the frontal faces in every frame of video sequence.

Input: M training samples $X = [x_1\ x_2\ \cdots\ x_M]$ which belong to c classes.
Output: Selected Haar functions v_i and their coefcients c_i.

Step 1: Compute the discriminant vectors $U = [u_1\ u_2\ \cdots\ u_{c-1}]$ of SRDA.

i) Compute the weight matrix W.

ii) Compute the eigenvector $v = [v_1,\ v_2,\ \cdots,\ v_M]^T$ of W.

iii) Calculate the vector u that satisfies $X^T u = v$.

Step 2: Approximate each $u_i \approx \tilde{u}_i = \sum_{j=0}^{K-1} c_j v_j$ using FHT.

i) Apply the FHT algorithm to compute the transformation coefficients c_i.

ii) Discard the small-valued coefcients and reserve the largest coefcients: $c_0 > c_2 > \cdots > c_{K-1}$.

iii) Obtain the approximation by $\tilde{u}_i = \sum_{i=0}^{K-1} c_i v_i$.

Figure 2.6: The training algorithm of FHT-SRD.

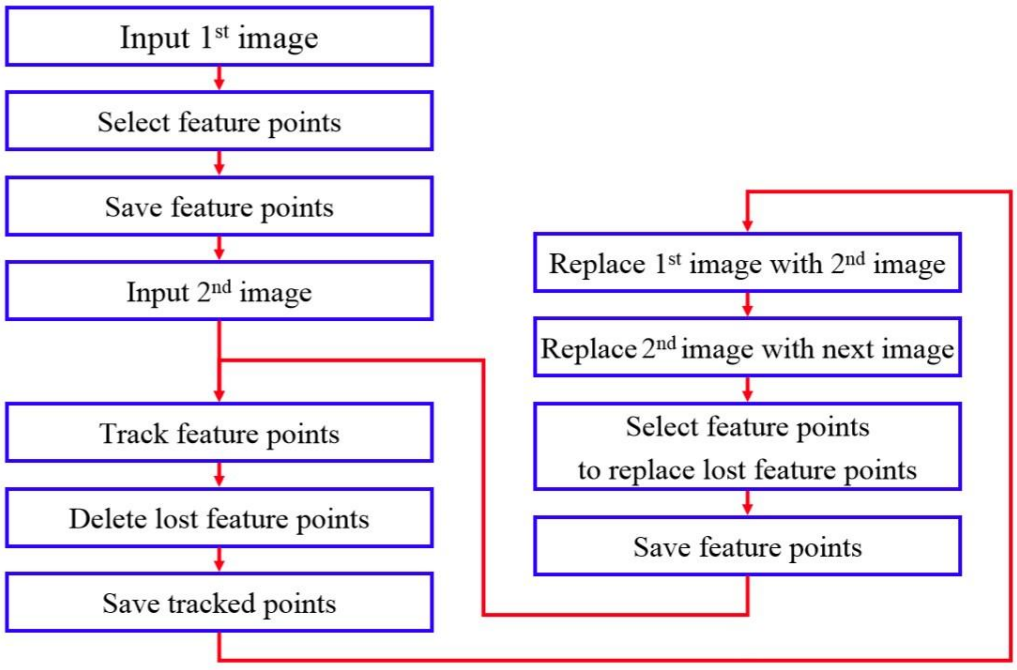


Figure 2.7: KLT-feature tracker algorithm.

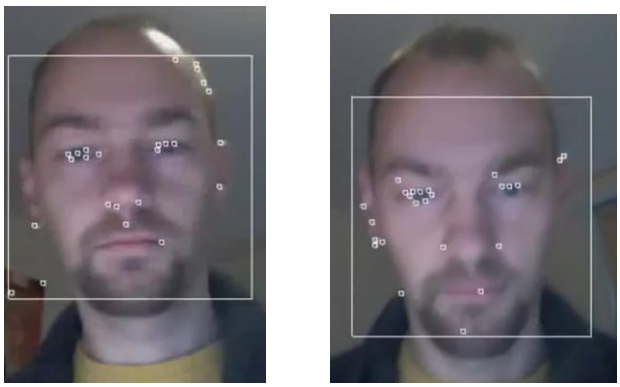


Figure 2.8: Extracting interest points using KLT-tracker

3. EXPERIMENTATION

To test the technique, a small demo program was implemented using MatLab version 8[9], on I5 PC with 2GB ram. The demo run on a video test and the image of the actor George Clooney was input as query image and the result shown in figures(3.1 and 3.2).

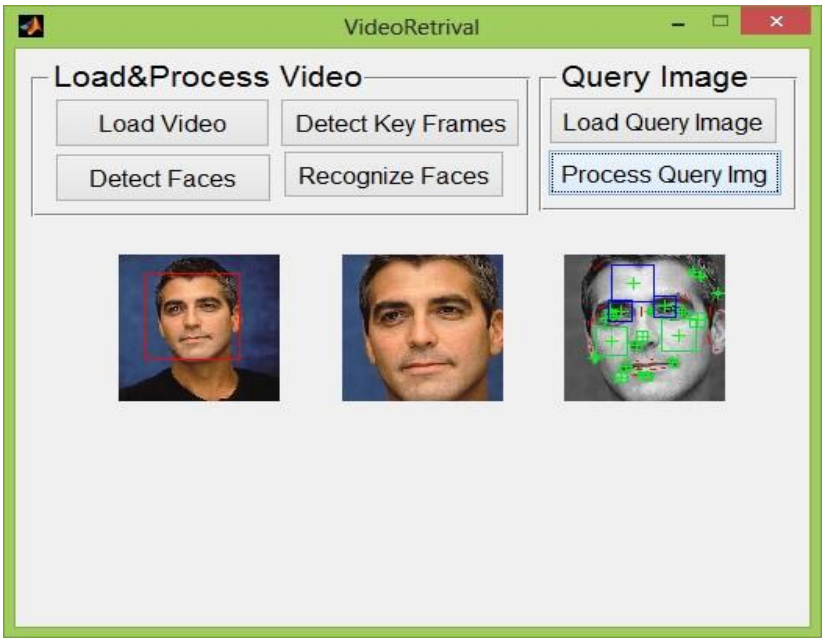


Figure 3.1: Tracking and feature extraction of the query face.

Figure 3.2: Retrieval of the similar faces.

4. CONCLUSION

This paper utilized Viola-Jones frontal face detector for detecting faces. 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 technique would be able to retrieve face frames from video on face image query, implementation done using Matlab.

5. REFERENCES

[1] J. Baber, "Video segmentation into scenes using entropy and SURF", IEEE 2011.
[2] Theo Ephraim, Tristan Himmelman, and Kaleem Siddiqi, "Real-Time ViolaJones Face Detection in a Web Browser", Canadian Conference on Computer and Robot Vision 2009.
[3] T. Nguyen et al., "An Efficient Method for Face Retrieval from Large Video Datasets", International Conference on Image and Video Retrieval (CIVR 2010), Xi'an, China, July 2010
[4] M. Yang, D. J. Kriegman and N. Ahuja, "Detecting Faces in Images: A Survey ", IEEE Transactions on Pattern Analysis and Machine Intelligence 2002.
[5] Nida Aslam, Irfanullah, K. K. Loo, and Roohullah " Limitation and ChallengesImage/Video Search & Retrieval", JDCTA 2009.
[6] Yanwei Pang et al., "Fast Haar Transform Based Feature Extraction for Face Representation and Recognition", IEEE Transaction on information forensics and security 2009.
[7] P. Viola, and M. Jones, "Robust Real-time Object Detection", International Journal of Computer Vision, 2001.
[8] Z. G. Sheikh, Thakare V. M. and Sherekar S. S., "Towards Retrieval of Human Face from Video Database: a Novel Framework", Journal of Information Systems and Communication 2012.
[9] "http://www.mathworks.com", Mathworks 2013.