

Video Based Face Extraction and Recognition

Hemant Patil¹, Richa Swami², Poorvi Tusam³

Fr, Conceicao Rodrigues College Of Engineering
University Of Mumbai, India

Abstract— This project implements face extraction as well as recognition. A video is provided as the input initially. All the faces from the video are extraction. These faces are further optimized to remove redundant data. A video is played resulting all the faces which were extracted from the input video. An image then is provided as an input for face recognition. This image is matched with the datasets of the faces extracted and if the match is found it is displayed as the result. Voila Jones algorithm is used for face detection. PCA algorithm is used for comparing frames. Eigen values are calculated which are used for comparison.

Keywords--Voila jones, PCA, eigen values.

Aim: The main aim of this project is to retrieve all the faces present in the given video stream. And to identify a face provided from the given video. This helps in automatically scanning and searching a face from a video without human interference.

I. INTRODUCTION

Video based face extraction and recognition in this project we have incorporated a method which helps user to extract all the faces from a video and store those faces as a video format. Input to this project is a video stream. First all the faces from a video are extracted and the required faces are recognized as per the query.

While traditional face recognition is typically based on still images, face recognition from video sequences has become popular recently as it provides more abundant information than still images. It is used in a wide range of applications that require real time verification of a person. For example, some applications involve checking for criminal records and detection of a criminal at a public place, finding missing people using the video streams received from the cameras fitted at public places, tracking down thieves, keeping a check of the unknown people entering a restricted area and many more.

Often, it gets difficult to employ face recognition in noisy videos with too many changes in poses and illumination. Besides, since video streams involve large amount of data, processing and detection of faces in real time becomes tough.

A. Objective

The objective of this system is:

- Face Extraction: From a stream of video, frames are generated which are then useful for detection and extraction of face. Video is also generated containing all the faces which were extracted.

- Face Recognition: All the extracted faces are further stored in a folder and are tested with image to be recognized. If match found it displays the image.

II. LITERATURE REVIEW

A. Adaptive Fusion of Multiple Matchers: [1]

In this system it was explored the adaptive use of multiple face matchers in order to enhance the performance of face recognition in video, and the possibility of appropriately populating the database (gallery) in order to succinctly capture intra class variations.

To extract the dynamic information in video, the facial poses in various frames are explicitly estimated using Active Appearance Model (AAM) and Factorization based 3D face reconstruction technique. It also estimates the motion blur using Discrete Cosine Transformation (DCT)

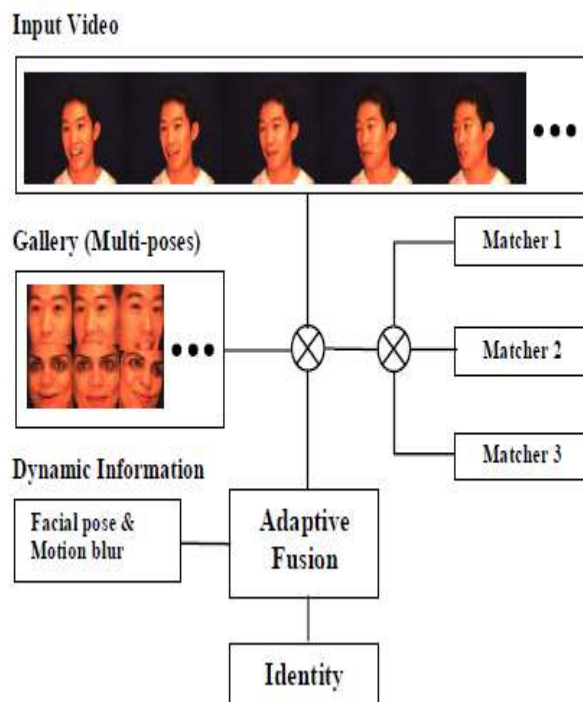


Fig 2.1. Identification using Matchers.

Flaws: In the system, Database size is very large due to multiple poses which increases the complexity of finding the true identity of a person. Due to multiple matchers the time taken to estimate the identity of the person is large.

B. Neural Network-Based Face Detection, 1998 [2]

By Henry A. Rowley, Shumeet Baluja and Takeo Kanade.

An image pyramid is calculated in order to detect faces at multiple scales. A fixed size sub-window is moved through each image in the pyramid. The content of a subwindow is corrected for non-uniform lightning and subjected to histogram equalization. The processed content is fed to several parallel neural networks that carry out the actual face detection. The outputs are combined using logical AND, thus reducing the amount of false detections. In its first form this algorithm also only detects frontal upright faces.

Flaws: It is not a motion based method, hence an additional scheme should be implemented to converted into still images first, and also face recognition isn't implemented only identification.

B. Frame Generation

First video is converted into frames. Normally it generates 30 frames/sec. Having 600 frames to process will take hours and unnecessary time for processing. That's why similar frames are eliminated using histogram. This is done by converting a frame(RBG) into gray scale image and generating its histogram. Difference of histograms is taken which are being compared and checked if they satisfy a threshold value. The histogram of images which satisfy the condition of the threshold are saved while others are discarded. This is used to eliminate redundant images.

Histogram: A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable. [3]

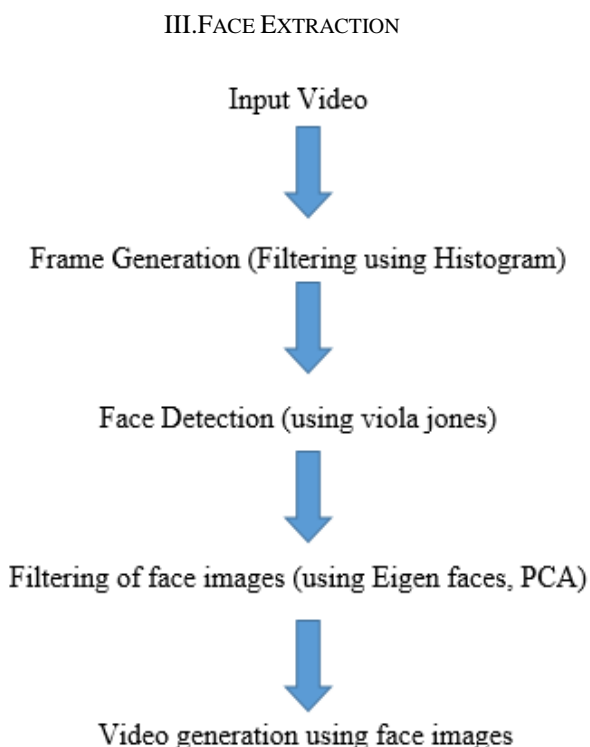


Fig.3.1 Flowchart of face extraction

A. Input Video

In this step we are passing a video of mp4,avi format which contains faces to be detected. The video also contains non-human figures. It can be a surveillance video of parking lot or shopping mall.



Fig.3.2 Input Image of a sunflower

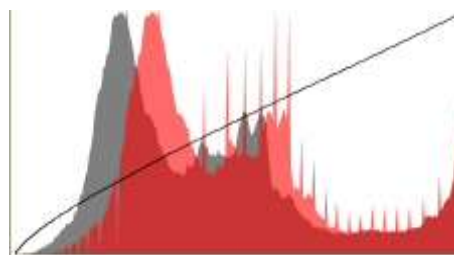


Fig.3.3 Histogram obtained of the input image of the sunflower

C. Face Detection

In this step faces are detected using cascade object detector. Cascade object detector uses viola jones to detects people's faces, eyes, nose, mouth and upper body.



Fig.3.4 Face detection from a frame

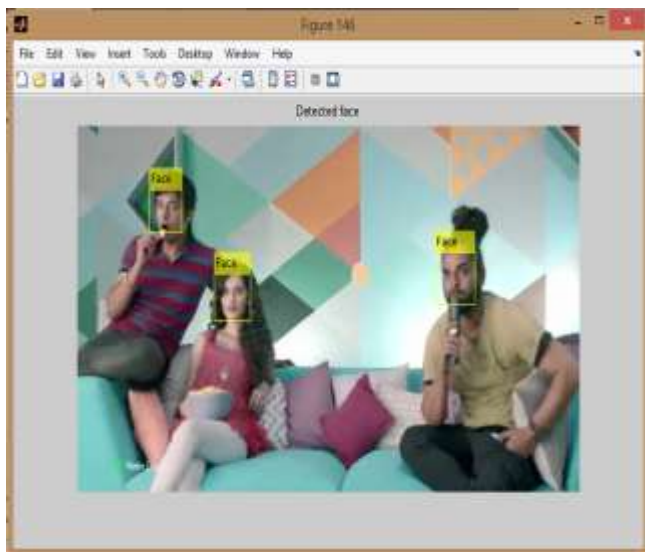


Fig.3.5 Number of faces detected in a frame

Viola Jones Algorithm : [2]

The algorithm has mainly four stages :

- Haar feature
- Creating integral image
- Adaboost Training algorithm
- Cascaded Classifiers

1) Haar feature: All human faces share some similar properties. This knowledge is used to construct certain features known as Haar Features.[6]

The properties that are similar for a human face are:

- The eyes region is darker than the upper-cheeks.
- The nose bridge region is brighter than the eyes.

That is useful domain knowledge:

- Location - Size: eyes & nose bridge region

- Value: darker / brighter

The four features applied in this algorithm are applied onto a face and shown below.

Rectangle features:

- Value = Σ (pixels in black area) - Σ (pixels in white area)
- Three types: two-, three-, four-rectangles, Viola & Jones used two-rectangle features



Fig.3.6 Haar Feature that looks similar to the eye region which is darker than the upper cheeks is applied onto a face

2) Integral image: 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. Because each rectangular area in a feature is always adjacent to at least one other rectangle, it follows that any two-rectangle feature can be computed in six array references, any three-rectangle feature in eight, and any four-rectangle feature in just ten. The speed with which features may be evaluated does not adequately compensate for their number, however. For example, in a standard 24x24 pixel sub-window, there are a total of possible features, and it would be prohibitively expensive to evaluate them all when testing an image. Thus, the object detection framework employs a variant of the learning algorithm AdaBoost to both select the best features and to train classifiers that use them. This algorithm constructs a “strong” classifier as a linear combination of weighted simple “weak” classifiers.

3) AdaBoost : AdaBoost (with decision trees as the weak learners) is often referred to as the best out-of-the-box classifier. When used with decision tree learning, information gathered at each stage of the AdaBoost algorithm about the relative 'hardness' of each training sample is fed into the tree growing algorithm such that later trees tend to focus on harder to classify examples.

4) Cascade Classifier : More the features lower will the false positive rate. But the larger the number of features greater will the time required for classification. Hence an optimized

framework is required. Generally an ideal optimization framework in which i) the number of classifier stages, ii) the number of stages in each stage, and iii) the threshold of each stage, are removed in order to minimize the expected number of evaluated features. In practice a very simple framework is used. Each stage in the cascade reduces the false positive rate and decreases the detection rate. A target is selected for the minimum reduction in false positives and the maximum decrease in detection. Each stage is trained by adding features until the target detection and false positives rates are met (these rates are determined by testing the detector on a validation set). Stages are added until the overall target for false positive and detection rate is met.[Viola jones 2]

D. Filtering of faces using PCA: [5]

First the PCA features of the face images are extracted:

- For this, the face image is first converted from RGB color format to GRAY color format using the function **rgb2gray(img)** in MATLAB.
- Next, the 2D image matrix is converted into 1D image vectors.
- A matrix **X** is created storing the image vectors of all images column wise.
- Next, the average face image vector **m** is computed and all the image vectors in **X** are subtracted with **m** to give matrix **A**.
- Matrix **A** is multiplied with its transpose to give a symmetric matrix **L**.
- Eigenvalues and eigenvectors of this matrix **L** are calculated
 $[V,D] = eig(L);$
 For eigenvalues > 1 the eigenvectors are stored in **L_eig_vec**.
- Finally the eigenfaces are calculated as
 $eigenfaces = A * L_eig_vec;$
- Projected image vector is calculated as
 $eigenfaces' * A(:,i);$
 A matrix **projectimg** is made adding the image vectors for all the images.

Next the PCA features of the test image are extracted:

- The face image is converted from 2D matrix to 1D image vector.
- The average face image vector **m (calculated earlier)** is subtracted from this image vector.
- Projection of face image onto the face space is calculated as
 $projtestimg = eigenfaces'*temp;$

Calculating & comparing the euclidian distance of all projected images from the projected test image:

- Euclidian distance is calculated as:
 $euclid_dist = (norm(projtestimg - projectimg(:,i)))^2;$
 for all images.

- A threshold is set as the standard deviation of the matrix **euclid_dist**
 $threshold = std2(euclid_dist)/2;$

First image is compared with the images ahead till a dissimilar image is found. The criteria for dissimilarity is the euclidian distance of the image being greater than the threshold. This image is then passed as the test image. This process continues till all the images have been processed in the same way. As a result we get filtered images.



Fig.3.7 Non optimized images



Fig.3.8 Optimized images

E. Video generation using face images[4]

In this step extracted faces are used to generate a video. The output video is in the AVI format. The output video contains only faces of people from input video. That makes user easy to identify people in surveillance video or any video. The output video is generated using a function known as “movie (h,M,n,fps,loc)”. Movie (h,M,n,fps,loc) specifies loc, a four-element location vector [x y 0 0], where the lower left corner of the movie frame is anchored (only the first two elements in the vector are used). The location is relative to the lower left corner of the figures or axes specified by handle h and in units of pixels, regardless of the object’s Units property.

IV. FACE RECOGNITION

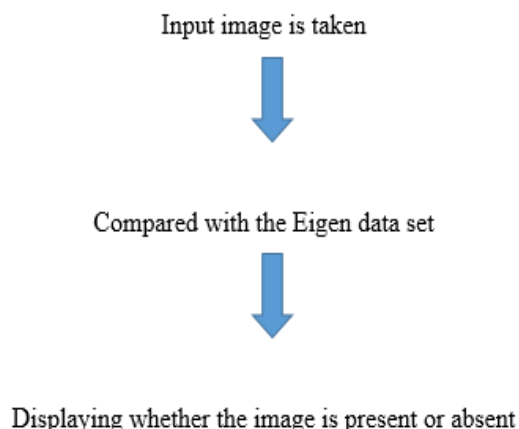


Fig.4.1 Flowchart for face recognition

The faces are recognized using the PCA algorithm, eigen vectors are used for comparing. In first step a test image is taken. In step two the Eigen vector of the image is calculated and compared with the Eigen vectors of the all stored images. If the match is found it shows the matched image otherwise it shows an error message showing “object not found”[5]

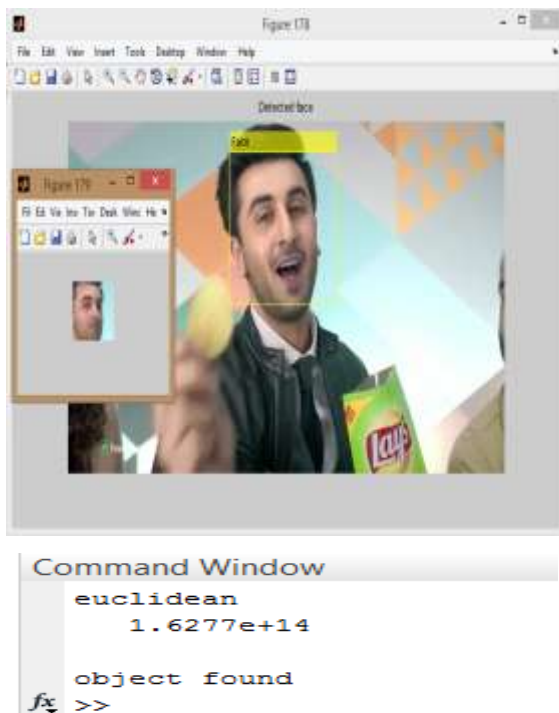


Fig.4.2 Displaying the output when the match is found for the test image in the dataset

V.APPLICATION

This system will be able to cater to the Security of the Customer. Companies providing security solutions can adopt this system to detect any culprits and suspects.

This system can be useful in crime investigations such as car thefts and other robberies.

The system allows not just detecting unknown faces but recognition all faces in the video.

This system will be used for surveillance at:

- >Shopping Malls
- >Commercial Parks
- >Housing Societies
- >General Parking Lots
- >ATMs

VI. SUMMARY

Video based face detection is used to detect face from a given input video. An image is passed, from these detected faces this test image is compared. If the match is found, the detected face is displayed from the dataset of the detected faces.

PCA algorithm is used for face extraction and recognition. In this, Eigen values are calculated for each image for comparison, in both face detection optimization, where similar redundant faces were ought to be eliminated and for face recognition where the Eigen value of the test image was being compared with the Eigen value of images in the dataset.

The Eigen face approach to face recognition was motivated by information theory, leading to the idea of basing face recognition on a small set of image features. The Eigen face approach provides a practical solution that is well fitted to the problem of face recognition. It is fast, relatively simple and has been shown to work well in a constrained environment. This method not only is helpful in face detection but also for face recognition.

VII. FUTURE SCOPE

The project can further be expanded and enhanced in following ways:

- The detected faces can be classified based on age, this will help us separate the young, middle aged and old people and narrow down the area of interest.
- Further, classification based on gender will also narrow down the area of interest.
- Classification of faces based on nationality will drastically bring down the number of faces of interest.
- Overall demographic information in that area can also be extracted from these detected faces for further research.

VIII. CONCLUSION

Hence we have extracted all the faces from a video, and even identified a specific face based on the image given for testing. Viola Jones algorithm is used for face detection i.e from the video all the faces are stored. We have used PCA for frame elimination and face recognition, where we have eliminated all the repeated faces and saved the rest. We use this dataset for further face recognition. Using this project we can successfully implement face extraction and recognition.

Earlier for identifying faces had to be done manually, which consumed a lot of time and efforts. It had high chances of faces being left undetected due to human errors. Hence motion based face extraction and recognition provides an efficient solution for extracting and identifying faces in a video. This method reduces the time consumption and helps in providing more accurate results.

IX. REFERENCES

[1] Unsang Park and Anil K. Jain. Michigan State University, East Lansing, MI 48824, USA (page no. 1)

[2] Implementing the Viola-Jones Face Detection Algorithm, Ole Helvig Jensen, Kongen Lyngby 2008 (page no. 9, page no.11)

[3] http://en.wikipedia.org/wiki/Image_histogram

[4] MATLAB Documentation :
<http://in.mathworks.com/help/matlab/ref/movie.html>

[5] Eigen faces for Recognition, Matthew Turk and Alex Pentland. Vision and modeling group, Themedia laboratory, Massachusetts institute of technology

[6]http://en.wikipedia.org/wiki/Viola%E2%80%93Jones_object_detection_framework

[7] A Fast Kernel Parametric Active Contour Method for Detecting Tumor based on Image Segmentation, Binil Kumar K ,Anoop B K, Adarsh K S