

# ***Criminals And Missing Children Identification Using Face Recognition And Web Scrapping***

S.AYYAPPAN

Department of Computer Science and Engineering,  
IFET College of Engineering,  
Villupuram, TamilNadu, India.  
Mobile No: +91-9025437794  
E Mail id: ajithayappan249@gmail.com

Dr.S.MATILDA

Department of Computer Science and Engineering,  
IFET College of Engineering,  
Villupuram, TamilNadu, India.  
Mobile No: +91-9442101456  
E Mail id: matildass@gmail.com

**Abstract**— Face recognition is a biometric based technology that maps an individual's facial features mathematically and stores the data as a face print. It employs Machine Learning on the image and generates a feature vector which maps an object with array of numbers. This technology is used by organizations such as Google and Facebook to create a digital profile for its users. This project proposes to use this technology for identifying criminals who are on the run from their previous records. An NCRB (National Crime Records Bureau) report shows that 70% of crimes are repeatedly committed by the same criminals. These criminals can be identified by the face recognition from an image or video frame which is captured by the cameras which are installed in various locations and it can also be used for identifying missing children. The disadvantage posed is that the images are usually blurred, have less clarity and not recognizable to the human eye. The proposed system can successfully recognize more than one face which is useful for quickly searching suspected persons as the computation time is very low. It creates a unique template for each face and compare them with other images available in dataset. If the match is found for the input face, then the details associated with the related image will be displayed. This system will decrease the crimes and ensure the security in our society.

**Keywords**— *feature extraction, machine learning, haar cascade classifier, web scrapping*

## I. INTRODUCTION

In recent years applications such as camera surveillance, healthcare etc., are achieved by using open computer vision technology for authentication, authorization, validation, and identification purposes. Web scrapping is a technique of extracting data and information from a website by accessing World Wide Web either directly through Hypertext Markup Language or by using a web browser. Python beautifulsoup is well known for web scrapping and it must use a parser to parse the formatted HTML text that are obtained from web pages. Most of the common facial recognition techniques include target matching method, geometric feature recognition method, and principal component analysis method and so on. Most of the criminals are mingled with us in our society and they are much hard to identify. The existing solutions for detecting these criminals are rely on huge datasets that are collected from various resources.

Traditionally, repeated criminals are identified by their biometrics such as thumbprint. But criminals are smart enough to not to leave their biometrics in crime scene. In developed countries, the government create dataset which is helpful for recognize the human face which compares the suspicious act with trained dataset and information stored in database.

Python-openCV can be used for detecting objects and also faces. But, they cannot identify rotational images. For detecting faces in rotated and blurry images, Haar Cascade Classifier is used.

Python Faceplib is used for extracting faces and it provides various functionalities to compare the features of two images. It is an API for recognize faces and comparing them. It takes two images as input and process them through different stages. The advantages of using faceplib are, it is able to detect faces in a blurry or long shot images. As the previous criminal's and missing children images require a small amount of preprocessing, they do not need to use haar classifier.

The web application is created by using python flask with jinja2 template and as it uses web scrapping no database or predefined dataset is needed for this application

Any changes made in information about criminal or a missing child will be dynamically updated. Face detection is done by using Haar cascade classifier. This classifier can detect at most 55 faces from an image. It converts the image from RGB to grayscale for face detection. The algorithm has four stages:

1. Haar Feature Selection
2. Creating Integral Images
3. Adaboost Training
4. Cascading Classifiers

Information about those criminals and missing children are collected from websites such as <https://www.fbi.gov/>, <https://www.interpol.int/> and so on.

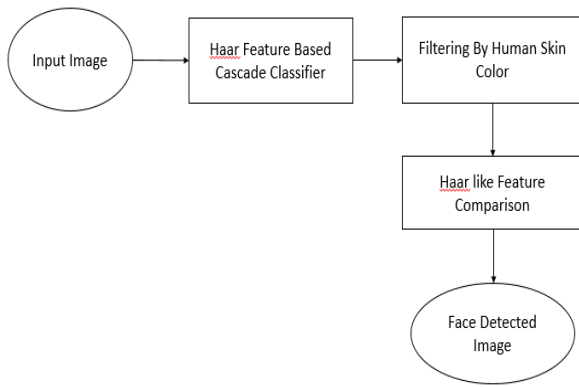


Fig 1: Face Detection using Haar Classifier

## II. LITERATURE SURVEY

Nurul Azma Abdullah presents an automated facial recognition system using known Principal Component Analysis approach for criminal database was proposed. That system is able to detect and recognize face automatically. This system will help law enforcement to identify the suspect in case of thumbprint is not in the scene. [1]

Apoorva and Impana discussed an approach to recognize faces using haar classifier and it uses single classifier for recognizing various kind of images with different qualities and rotations. It uses several weak classifiers instead of a complex one. [2]

Kavushica Rasanayagam explains face detection and recognition with deep learning approach based on Convolution Neural Network (CNN) technique. It classifies data by using IMDB dataset with AWS cloud. [3]

Teddy Mantoro reviews eigenface methods to create a face vector or face print by using cascade classifiers and uses Principal Component Analysis (PCA) for training. [4]

Liping Chang describes pattern recognition for classification in feature extraction based on a stacked convolutional auto encoder (SCAE) with deep learning. Further, it combines SRC with Local Binary Projection (LBP) features. [5]

MING Ju-wang exposes a 3D approach for choosing faces by simulating based on 2D images. It can also detect multiple faces dynamically with varying situations. [6]

Mohd Yusuf Firoz Siddiqui explains about training the faces efficiently based on symmetrical and mirror images by reflecting the original face using Principal Component Analysis and some fusion techniques. [7]

Hyung-II Kim describes about face image assessment of objective visual quality with the improvement in quality of images for training a dataset in a convenient and reliable way.

It extracts a way of automatic face recognition and the improvement in face image quality (FIQ) [8]

Piyush Kakkar explains detecting faces from video and normalize the picture quality that is extracted, recognize and compare the faces. It uses so many datasets from different countries which leads to a bottleneck in performance. [9]

Lamiaa A. Elrefaei, Alaa Alharthi, Huda Alamoudi, and Shatha Almutairi presents a client-server approach for real-time tracking and it uses optical flow networks that is implemented in android lollipop version. It extracts features in an image such as corner features and regular features .It uses Adaptive Mean Algorithm for the camera based surveillance. [10]

## III. EXISTING SYSTEM

The most common way by which criminals are identified in our country is using their thumbprint or other biometrics. The architecture of existing approach is illustrated in Fig 2. It involves four major steps. They are,

### A. Image Acquisition

The images of suspects are given as input. The images are read as a numpy array and stored using opencv. The criminal and missing children images are acquired from predefined datasets that is collected from local and international cop departments.

Feature	Value
Face color (red)	234
Face color (green)	192
Face color (blue)	134
Width	16
Height	23
Hair color (red)	044
Hair color (green)	034
Hair color (blue)	043
Spectacles	1
Facial hair	0

Table 1: Feature Extraction from Face Images

## B. Face Detection

Python opencv is used to detect faces and it usually uses principal component analysis [4] or convolution neural network (CNN) [3] technique. These techniques are used to classify images as face and non-face regions based on features obtained from an image. For such a task the classifier must be trained with positive and negative face images to identify the faces in the given image.

## C. Preprocessing

The detected image is cropped and stored in the form of a numpy array. The detected image is preprocessed in order to remove the unwanted features to reduce the amount of computation performed. It provides a 3D view of the face that is stimulated from the face image. It fills up the regions on face automatically by applying deep convolutional networks.

## D. Feature Extraction

It extracts the features of the face such as the color, width and height of the face, color of hair, spectacles, and facial hair and so on by employing deep learning. These features can be represented as an array of numbers which will be unique for all the faces and they are stored in JSON format. Table 1 explains some of the features mapped to a value of integers.

## E. Template Matching

The extracted features of the face are called as a face print or feature vector or simply a template. It will compare the two faces by the similarities of their features. Based on the similarities the confidence level of matching is calculated. If the confidence level crosses the threshold value, then the image will be considered as matching one. The threshold value is fixed manually by choosing the best value of matching which increases the accuracy level of the comparison by analyzing the results. Based upon the result of template matching the final decision is made to determine whether the suspect is criminal or not.

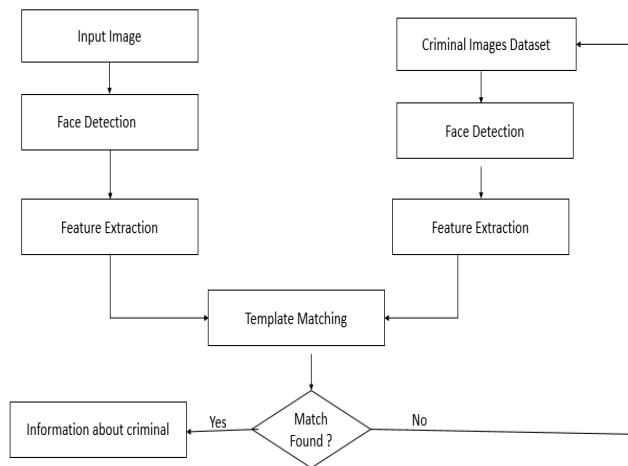


Fig 2: Flow diagram for face recognition

## DRAWBACKS

The main disadvantage of existing method is it is completely rely on a static dataset and requires a prolong time interval for training the face database. For instance if a missed child is found, then it should be changed in dataset to avoid anomalies. But, any small change made in the dataset or database requires high training time which degrades the performance and sometimes it leads to an inaccurate result.

## IV. PROPOSED SYSTEM

The work done in our proposed system is explained in Fig 3. It is consisting of following modules:

### A. Image Acquisition

Here, the suspect image is provided by user via a user interface and the previous criminal and missing children images are scrapped from web. It can access the sites dynamically so that the updated information will always be provided. It can also use videos from camera or CCTV footage. Instead of reading videos frame by frame, it reads in seconds to avoid redundant images as the videos have 29FPS.

### B. Face Detection

The face detection is done by using the Haar Cascade Classifier.

#### Haar Cascade Classifier

It is an approach based on machine learning that uses a cascade function which is trained from a bunch of positive and negative images. It is can also be used to detect objects in other images.

$$F(\text{Haar}) = \sum F_{\text{White}} - \sum F_{\text{Black}}$$

Where:

$\sum F_{\text{Black}}$  = sum of pixels of the dark area

$\sum F_{\text{White}}$  = sum of pixels of the bright area

$F(\text{Haar})$  = the Haar-like feature

### C. Scraping Web

Python requests library is used along with BeautifulSoup for scrapping images from web. The sites containing information about criminal and missing person are specified and the source code from the site is obtained from BeautifulSoup. The Html5lib parser is used for obtaining unformatted text as an HTML document.

The required elements are obtained by specifying the tag and class name in which it is consisting of and process them based on the type of information. This information are dynamically updated and it doesn't require additional training for little changes.

#### D. Feature Extraction

The features from an image is obtained by using opencv with detect multiscale and the parameters should mentioned with correct values to extract the features. It employs more than one classifier on obtained face images. The image is converted from RGB to Grayscale and HSV colors with modified XML of cascade classifiers. The extracted features stored as an array of integers with unique value for each feature. The mapping is done by the use of deep learning approach.

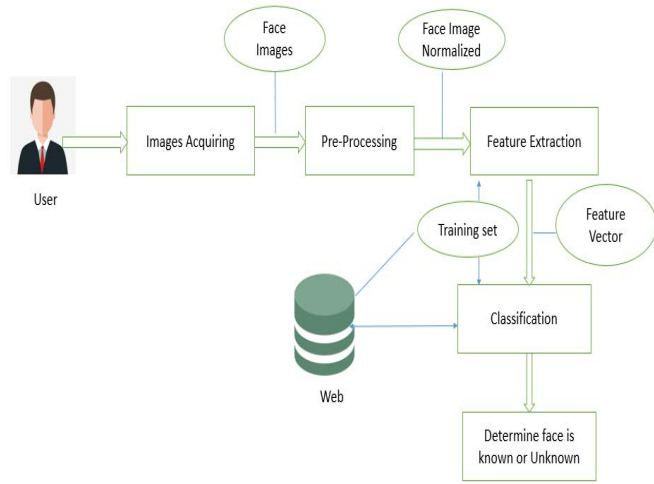


Fig 3: Face comparing with web images

#### E. Comparing Templates

The face vectors are compared for their similarity. The similarity level is calculated based on the matching features of both images. For comparing the templates facepplib API library of python is used. It takes faceprint from suspect input image and the original image of the criminals. Then, it produces faceprint for criminal images and compare them.

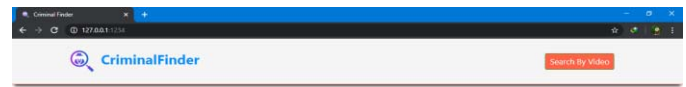
##### Python Facepplib

Python is a powerful and well documented language that can be used for various applications. On computer vision technology, python is widely used for its simplicity. Facepplib API provides an easy to perform the template matching. The configurations for it is done by api\_key and secret\_key configuration.

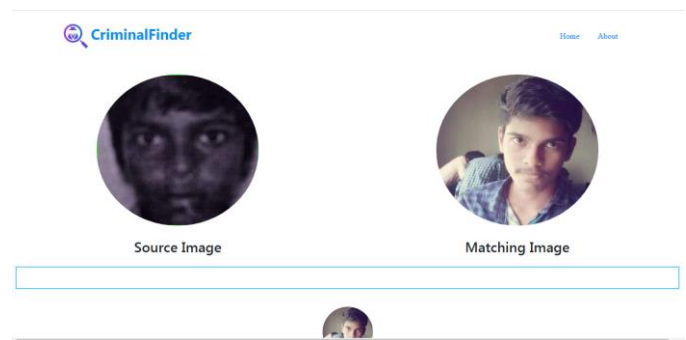
It provides compare method which returns all the similarity features of two templates. The confidence level calculated based on distance between two features. If two features are near then they are highly likely to be same. Every set of faces has a confidence value as some of the features may be similar both the faces. By accessing the confidence value for different faces in analysis a threshold value is provided. If the confidence value is higher than threshold value, then the two faces are considered as the same and the person is decided as a criminal or a missing one. The respective information available about the face image in web is displayed.

#### V. RESULTS

The system is tested with various types of images. For experimental purpose it is tested with a sample dataset for analyzing its accuracy level with various images such as blurred, long shot, childhood images and aadhar card images. It takes the input image and using flask application it performs comparison of images with web images and display the results which enables is to decide whether the person in image is a criminal or not.



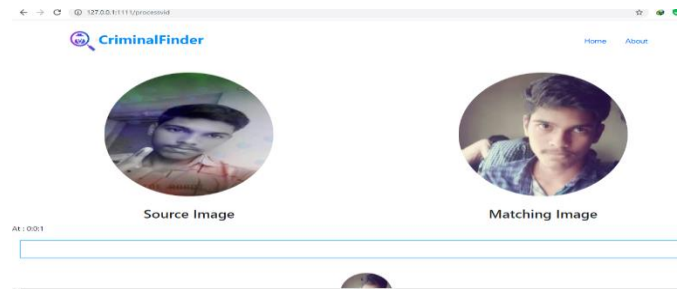
##### From Aadhar Card Images:



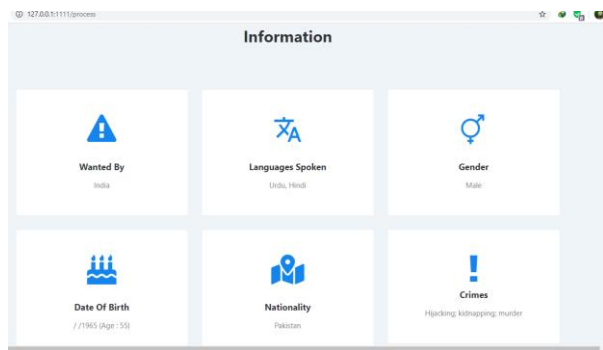
##### From Childhood Images:



##### From Videos:



### Displaying Information:



- [9] Piyush Kakkar, Mr. Vibhor Sharma (2018) "Criminal Identification System Using Face Detection and Recognition". International Journal of Advanced Research in Computer and Communication Engineering
- [10] Lamiaa A. Elrefaei, Alaa Alharthi, Huda Alamoudi, Shatha Almutairi (2017) "Real-time Face Detection and Tracking on Mobile Phones for Criminal Detection".

### CONCLUSION

In this work, we compare the various types of images and the accuracy level of results is very satisfying. It performs well with both images and videos. The results displayed are 90% accurate. This requires less memory space to implement and takes less time when compared with other approaches. By using this the criminals and missing children/person can be easily identifiable and it keeps on updating dynamically. The analysis process carried out with real criminal images in the web and it provides good results. We believe that, this application will decrease the crimes in our environment.

### REFERENCES

- [1] Nurul Azma Abdullah, Md. Jamri Saidi and Nurul Hidayah Ab Rahman "Face recognition for criminal identification: An implementation of principal component analysis for face recognition" The 2nd International Conference on Applied Science and Technology 2017 (ICAST'17)
- [2] Apoorva.P, Ramesh.B and Varshitha.M.R "Automated criminal identification by face recognition using open computer vision classifiers" Third International Conference on Computing Methodologies and Communication (ICCMC 2019).
- [3] Rasanayagam, K.Kumarasiri, S.D.D, Tharuka, W. A. D. Samaranayake, N. Samarasinghe and P. Siriwardana "CIS: An Automated Criminal Identification System". 2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS) R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [4] Mantoro, T., Ayu, M. A., & Suhendi. (2018). "Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods" 2018 6th International Conference on Multimedia Computing and Systems (ICMCS).
- [5] Chang L, Yang J, Li S, Xu H, Liu K & Huan, C. (2018). "Face Recognition Based on Stacked Convolutional Autoencoder and Sparse Representation". 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP).
- [6] MING Ju-wang (2018), "Face Feature Dynamic Recognition Method Based on Intelligent Image". International Conference on Virtual Reality and Intelligent Systems
- [7] Mohd Yusuf Firoz Siddiqui and Sukesha (2015), "Face Recognition using Original and Symmetrical Face Images". 1st International Conference on Next Generation Computing Technologies (NGCT-2015)
- [8] Hyung-Il Kim, Seung Ho Lee, and Yong Man R (2015), "Face Image Assessment Learned With Objective and Relative Face Image Qualities for Improved Face Recognition"