International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Intelligent CCTV Footage Search System

Akhilesh Shinde[1], Aniket Shinde[2], Vivek Singh[3] , Priya Tiwari[4] , Prashant Lokhande[5]

[1,2,3,4]*Student, Dept .of Computer Engineering, Pillai College of Engineering, Maharashtra, India*
[5]*Professor, Dept. of Computer Engineering, Pillai College of Engineering, Maharashtra, India*

Abstract - *At present law and enforcement agencies mostly depend on CCTV footage to trace the criminal. Sometimes, crime involves searching CCTV footage of various locations of numerous days and duration. This will pose the biggest hurdle in the investigation process as it involves tremendous manpower to search the suspect in CCTV footage. This project proposes intelligent searching of CCTV footage with an added module of Machine Learning. To build an Intelligent Search security system which takes suspect image as an input and searches the suspect in CCTV footage by considering the parameter such as appearance of suspect and facial orientation features using machine learning. The system will return the position of suspect in CCTV footage video reel with highlighted suspect figure time when the suspect was identified in CCTV footage. Various algorithms are reviewed and planned to use the following algorithms like Haar cascade for face detection and CNN and one shot learning method for face recognition and comparison.*

Key Words: **Neural network, suspect, Face Detection, MTCNN, Haar Cascade, Face Recognition, CNN, CCTV footage**

1. INTRODUCTION

Nowadays CCTV poses an important role in security. They are helpful in surveillance of public and private properties. Also CCTV helps the law and enforcement department to solve the crimes. It provides strong evidence in the form of video footage of actual crime and criminals. Thus suspects of crime can be easily identified and based on this police can search and trace the suspect. But searching a particular suspect in a CCTV footage having hours of duration is a very tedious job. It requires manpower and is a time consuming process. Intelligent CCTV footage search system provides a solution for this by searching suspect images in CCTV footage using Face detection and Face recognition technologies.

In this project CCTV footage and suspect image will be taken as input and output will be a video frame that consists of suspect involvement. For this Convolutional neural networks are used for face detection and recognition. Preprocessing of video footage and suspect image is done using various image filtering and enhancement techniques.

2. LITERATURE REVIEW

A. Deep Neural Network

This technique is used by Dr. Priya Gupta et al.[1]. Deep Neural Network is feed forward ANN with multiple hidden layers and higher layers of abstraction. The width of the Deep neural network is determined by the dimensionality of the hidden layer. This type of neural network also consists of a dropout layer which drops or eliminates the nodes between hidden layers to reduce the complexity. The input of this neural network is the extracted features of faces while processing of face image is done beforehand i.e raw data input is not fed to the neural network. [1]

B. Accumulative Differences Images (ADI)

This technique is developed by Author -Ade Nurhopipah et al.[2]. Here the focus was developing a motion detection system. In this research, motion segmentation is carried on using ADI. In this method, not only one but some images are compared to references. Comparison results between the referent images with tested images will be accumulated and compared to a certain threshold. For example, f(x,y,t1) = R(x,y) is the referent image and f(x,y,tk)= f(x,y,k) is the k–image where k>2, therefore, ADI value can be defined in equation,

$$A_k(x,y) = \begin{cases} A_{k-1}(x,y) + 1, & if\ |R(x,y) - f(x,y,k)| > T_A \\ A_{k-1}(x,y), & otherwise \end{cases}$$

Next, determine the threshold which sets the accumulation limit. This limit determines the presence of motion. If it exceeds the threshold, the image has motion. [2]

C. Cascade method for merging classifiers

After ADI Cascade method for merging classifiers is used for detection. Cascade structure increases the speed of the detector by focusing on areas where there are objects in the image. The best features will form a strong classifier which classifies faces into positive and negative images. Here, a classifier is applied in order to reject the non-face sub-window. The whole process of this detection process is aimed to grow a decision tree called cascade as shown in Figure [2]

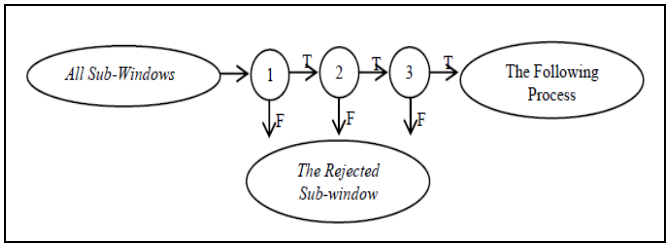

Fig -1: Detection scheme with cascade classification [2]

D. Adaptive Principal Component Analysis

The Adaptive principal component analysis is discussed by Ting Shan et al.[3]. Adaptive Principal Component Analysis (APCA) inherits characteristics from both PCA and Fisher Linear Discriminant by warping the face subspace according to the within- and between-class covariance. In the text below we shall use the following notation: sj,k indicates a feature vector for the j^{th} class and k^{th} sample from that class sj,0 represents the vector for the reference image in the j^{th} class, si, j,k indicates the the i^{th} element of the vector sj,k, Kj indicates the number of samples in class j and N is the number of classes. APCA is comprised of three steps, which are briefly overviewed below

. • Subspace Projection, where standard PCA is used to project every face image into the face subspace to generate m-dimensional feature vectors sj,k.

• Whitening Transformation, where the subspace is whitened according to its eigen values, with a whitening power p. This compensates for the overweighting of leading eigenvectors. The corresponding transformation matrix is:

$$\mathbf{C}_{cov} = \text{diag}\left[\lambda_1^{-2p}, \lambda_2^{-2p}, \cdots, \lambda_m^{-2p}\right]$$

• Filtering, where the feature vector elements are weighted according to the Identification to Variation value ψ with a filtering power q. The corresponding transformation matrix is: [3]

$$\mathbf{R} = \text{diag}\left[\psi_1^q, \psi_2^q, \cdots, \psi_m^q\right]$$

E. Neural Network for feature extraction

In paper published by Vinodpuri Rampuri Gosavi et al.[4] have discussed several methods for facial feature extraction techniques. Such as Geometry based face face extraction, Principal component analysis, Convolutional neural network and Back propagation network. [4]

F. Convolutional Neural Networks:

This technique is modelled by Shraddha Arya et al.[5]. In neural networks, Convolutional neural networks (ConvNets or CNNs) is one of the main categories to do image recognition, image classifications. Objects detections, recognition faces etc, are some of the areas where CNNs are widely used. Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernels), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. [5]

G. High-Fidelity Pose and Expression Normalization

This method is discussed by Xiangyu Zhu et al.[6]. In this paper, the method based on pose and expression normalization is presented which builds an algorithm to recover a canonical view and expression free image with high fidelity. In the process, firstly face landmark alignment is done which is then further processed with 3DMM that is fitted into the original image that goes on for further identify transformation by meshing the whole image into 3d object and normalizing it with 3D transformation. Trend fitting and detail filling method is used to fill in the invisible region leading to smooth results.[6]

H. Multi-task Cascaded Convolutional neural network (MTCNN)

This method is discussed by Kaipeng Zhang et.al [7] particularly for predicting face and landmark location in a coarse-to-fine manner and for joint face detection and alignment..Our method achieves superior accuracy over the state-of-the-art techniques on the challenging FDDB and WIDER FACE benchmark for face detection, and AFLW benchmark for face alignment, while keeps real time performance.[7]

I. Neural Networks:

Artificial Neural Networks (ANN) is technology in the Machine learning domain that consists of a series of algorithms that understand the relationship between the dataset in a way any human brain works. Advantage of neural is unlike traditional computer algorithms it is adaptive in nature. It learns from training dataset and performs prediction on further test dataset. It consists of an input layer, output layer and intermediate hidden layers. Each input is fed with some weight associated with each input. These weights depict the importance of inputs to produce the final result. The mapping of input to output is done using some activation function. Many researchers have used this method in the facial detection and recognition field. Some of them are Deep Neural Network and Convolutional Neural Network. [8]

2.1 Summary of related work

Table -1: Summary of Literature review

Literature	Haar Cascade	Adaptive PCA	Hybrid
Dr. Priya Gupta et al. 2018 [1]	Yes	No	No
Ade Nurhopipah et al. 2017 [2]	Yes	Yes	Yes
Ting Shan et al. 2017 [3]	No	Yes	Yes
Dr. G. S. Sable et al. 2018 [4]	No	Yes	Yes

3. PROPOSED WORK

The proposed system given with a suspect image will perform intelligent searches on CCTV footage to find the suspect. The proposed system will accept CCTV video file and suspect image as the input and provide output file as video frame and time stamp in which suspect is being spotted. The video quality of CCTV footage can be of low resolution and the image of the suspect may need to be enhanced.

The approach of the proposed system begins by converting input CCTV video file into video frame per seconds using OpenCV library of computer vision. The enhancing and up scaling of suspect images is done using various image processing techniques. Some of the techniques are filtering with morphological operators, histogram equalization, noise removal using wiener filter, median filtering, etc. These image processing techniques are applied to suspects using different machine learning methods such as bicubic interpolation. The next phase is face detection in the video frame. The face is detected in video frames using facial machine learning algorithms i.e Haar Cascade Classifier method. Once the face is detected in the video frame as per the given threshold value then only that frame is considered for further processing and frames without any faces are rejected. The features of detected faces are extracted and stored using a convolutional neural network model. Similarly facial features of suspects are extracted and stored.

Next phase is comparison of faces present in each frame with the suspect face. To carry out the face recognition we are proposing one shot learning technique. It is a classification task where one, or a few, examples are used to classify many new examples in the future. Modern face recognition systems approach the problem of one-shot learning via face recognition by learning a rich low-dimensional feature representation, called a face embedding, that can be calculated for faces easily and compared for verification and identification tasks. In other words, the training dataset will be only one image of each person and learn from that one example to recognize that person again. Similarity function is defined to calculate similarity between suspect face and faces present in each frame of video footage. The frame containing the most similar face with a suspect face is stored as output along with the timestamp of the frame.

3.1 System Architecture

The system architecture is given in Figure 2. Each block is described in this section.

A. Input

The system will be given CCTV video footage and suspect image file as input. Certain image enhancement techniques could be used in order to enhance the CCTV footage as well as suspect image. The input video footage will be converted to video frames.

B. Face Detection

In order to detect the faces from suspect image and video frames haar cascade algorithm is applied where haar cascade is a machine learning object detection algorithm used to identify objects in an image or video and based on the concept of features proposed by Paul Viola and Michael Jones in their paper "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001. It uses Haar features Haar Cascade classifier is based on the Haar Wavelet technique to analyse pixels in the image into squares by function. This uses “integral image” concepts to compute the “features” detected. Haar Cascades use the Ada-boost learning algorithm which selects a small number of important features from a large set to give an efficient result of classifiers then use cascading techniques to detect faces in an image. The haar features used are edge features, line feature, rectangular feature. The face is detected and cropped from frames and suspect image.

C. Facial feature extraction

The features of faces are extracted using VGG facenet CNN model. FaceNet was introduced in 2015 by Google researchers. It transforms the face into 128D Euclidean space similar to word embedding. Once the FaceNet model has been trained with triplet loss for different classes of faces to capture the similarities and differences between them, the 128 dimensional embedding returned by the FaceNet model can be used to cluster faces effectively. In this method the facial features are transformed into 128D face embedding(vector space) further stored in the database. The model is trained with the face of the suspect to create its

128D face embedding as well as for the faces detected in each video frame.

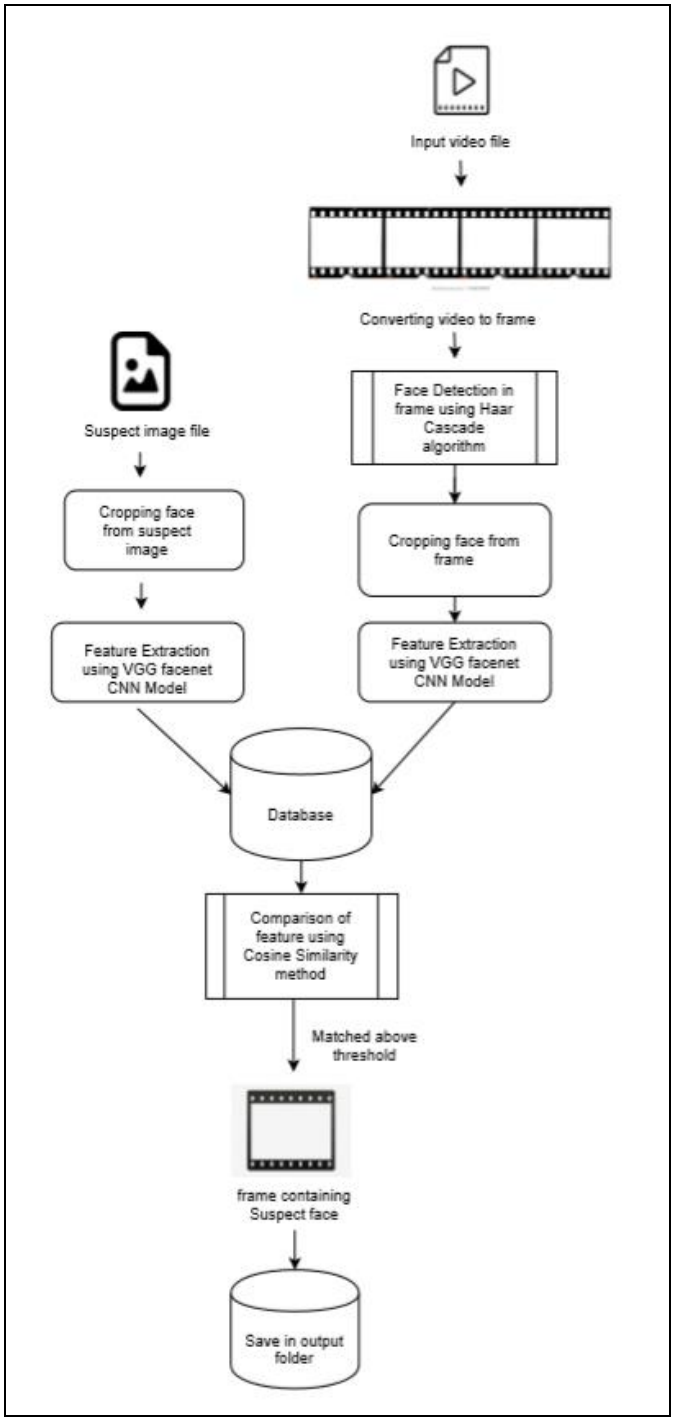

Fig -2: Proposed system architecture

D. Comparison of features

The facial recognition purpose is achieved using the one shot learning approach. One-shot learning is an object categorization problem in computer vision. Whereas most machine learning based object categorization algorithms require training on hundreds or thousands of images and very large datasets, one-shot learning aims to learn information about object categories from one, or only a few, training images. One-shot learning can be implemented using a Siamese network. This network has got two identical fully connected CNNs with the same weights and accepting two different images. Normal CNN uses softmax to get the classification, but here the output of a fully connected layer is regarded as 128 dimensional encoding of the input image. First network output the encoding of the suspect input image and second network output the encoding of each face obtained in video frames. Finally, we can say these encodings are the good representation of these input images. These two faces are compared by calculating distance between them. For this purpose cosine similarity method is used. In a way, distance would be closer for similar faces and further away for non-similar faces. Certain threshold is fixed above which if the face is matched then that frame is stored as output.

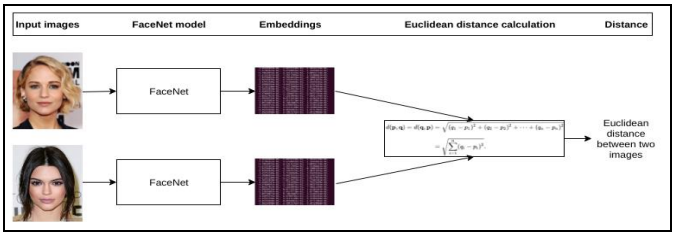

Fig -3: Face comparison using one shot learning [14]

E. Output

It is the last phase of the proposed system, it will contain the final output of the system. The output file contains a video frame in which the suspect is spotted and a timestamp of the video frame.

4. REQUIREMENT ANALYSIS

Different face recognition methods have been analyzed in this paper. Eigen faces is a facial recognition technique which works on unsupervised dimensionality reduction technique called principal component analysis (PCA). Dimensionality reduction is a type of unsupervised learning where we want to take higher-dimensional data, like images, and represent them in a lower-dimensional space. Using eigenfaces one can achieve accuracy over 90% accuracy. But requires a large database of faces.

Another approach is using neural networks. Machine learning is proved efficient in the facial recognition field. Convolutional neural networks are the most popular network used for facial recognition. Face images are trained through multiple layers of neural networks and then tested for recognizing the face. But in order to accurately identify the face, the training dataset has to be large enough.

Also certain geometric based methods are also present in this field such as SVM [Support Vector Machines], PCA

[Principal Component Analysis], LDA [Linear Discriminant Analysis], Kernel methods or Trace Transforms. But they also have limitations in terms of performance and accuracy.

The problem defined here consists of only one or two face images of the suspect which contradicts the requirements of above mentioned methods. So the paper proposes the one shot learning approach which satisfies the requirement of a problem of limited dataset. One-shot learning is an object categorization problem in computer vision. In this a CNN model is trained only with one image and then classified further from other images using cosine similarity. Firstly the facial features are encoded to 128D face embedding(vector space). Then the distance between two faces is calculated using similarity methods. The distance between two similar faces are very less whereas two dissimilar faces have a large distance. In order to obtain the output a certain threshold is set which must be satisfied if the suspect face is matched with any of faces in the video frame. The frame containing the suspect satisfying the threshold condition is stored as output along with the timestamp.

5. CONCLUSION

Face recognition is a widely used phenomenon and it can be used in security and surveillance systems in order to track the suspect in CCTV footage. CCTV footage search system moves a step forward in finding out the suspect in the CCTV footage. This system makes use of a pre-trained network which is used for feature extraction from a single image. The system consists of CNN networks which represent feature embeddings of query image. Features here are extracted at the highest layer. During the testing time, the system is provided with a suspect image which passes through the same network and compares its features extracted with the features that were extracted during the time of training. The system learns a similarity function known as cosine similarity which is used for comparison. This paper presents a flexible approach of one shot learning by using minimum data to train a face recognition model.

REFERENCES

[1] Dr. Priya Gupta, Nidhi Saxena, Meetika Sharma, Jagriti Tripathi, "Deep Neural Network for Human Face Recognition" (2018)

[2] Ade Nurhopipah, Agus Harjoko, "Motion Detection and Face Recognition For CCTV Surveillance System" (2018)

[3] Ting Shan, Shaokang Chen, Conrad Sanderson, Brian C. Lovell, "Towards robust face recognition for intelligent-CCTV based surveillance using one gallery image" (2007)

[4] Dr. G. S. Sable, Dr. Anil K Deshmane, Vinodpuri Rampuri Gosavi, "Evaluation of Feature Extraction Techniques using Neural Network as a Classifier : A Comparative Review for face Recognition" (2018)

[5] Shraddha Arya, Arpit Agrawal, "Face Recognition with Partial Face Recognition and Convolutional Neural Network" (2018)

[6] Xiangyu Zhu, Zhen Lei, Junjie Yan,Dong Yi, Stan Z. Li, "High-Fidelity Pose and Expression Normalization for Face Recognition in the Wild" (2015)

[7] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Senior Member, IEEE, and Yu Qiao, Senior Member, IEEE, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks" (2016)

[8] Debaditya Acharya, Kourosh Khoshelham, Stephan Winter, "Real-time detection and tracking of pedestrians in CCTV images using a deep convolutional neural network" (2017)

[9] Neel Borkar, Sonia Kuwelkar,"Face Recognition System Using PCA, LDA & Jacobi Method"(2017)

[10] J.Chen, "Neural Network", 2019. [Online]. Available: https://www.investopedia.com/terms/n/neuralnetwork.asp

[11] "Bicublic interpolation", 2019. [Online]. Available: https://en.wikipedia.org/wiki/Bicubic_interpolation

[12] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way", 2018. [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53\

[13] W. Berger, "Deep Learning Haar Cascade Explained", 2019. [Online]. Available: http://www.willberger.org/cascade-haar-explained/

[14] Dhanoop karunakaran, "One Shot learning explained using facenet", 2018. [Online]. Available: https://medium.com/intro-to-artificial-intelligence/one-shot-learning-explained-using-facenet-dff5ad52bd38

[15] Satyam Kumar, "Haar Cascade face identification", 2019. [Online]. Available: https://medium.com/@krsatyam1996/haar-cascade-face-identification-aa4b8bc79478