2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)

Proposed System for Criminal Detection and Recognition on CCTV Data Using Cloud and Machine Learning

Samit Shirsat
Computer Engineering Department
S.R.I.E.I.T,
Shiroda-Goa
samitshirsat@outlook.com

Aakash Naik
Computer Engineering Department
S.R.I.E.I.T,
Shiroda-Goa
naikaakash939@@gmail.com

Darshan Tamse
Computer Engineering Department
S.R.I.E.I.T,
Shiroda-Goa
darshan.tamse@gmail.com

Jaysingh Yadav
Computer Engineering Department
S.R.I.E.I.T,
Shiroda-Goa
jaysinghyadav9197@gmail.com

Pratiksha Shetgaonkar
Computer Engineering Department
S.R.I.E.I.T,
Shiroda-Goa
pratiksha.gawas@gmail.com

Shailendra Aswale
Computer Engineering Department
S.R.I.E.I.T,
Shiroda-Goa
aswale.shailendra@gmail.com

Abstract— **Finding criminals or hunting for people, in a CCTV video footage, after a crime scene or major attack takes place, is a time consuming task. As informed to us by cyber cell members of Goa branch, they make multiple members of the department sit with laptops and computers literally to search through the CCTV footage to find and trace the guilty, as they don't have the automated system for doing this task with them. This process is both time and labor intensive.**
In this research paper we have tried to survey the existing technologies as well as we propose a new system for criminal Detection & Recognition using Cloud Computing and Machine Learning, which if used by our Crime Agencies would definitely help them to find criminals from CCTV footage. The proposed system can not only help find criminals but if used properly on different sites such as railway stations etc, can also help find missing children and people from the CCTV footage available from the respective site.
Existing solutions use traditional face recognition algorithms which can be troublesome in changing Indian environments especially factors like light, weather and especially orientation. Some CCTV are in a bad place and can get tilted resulting in a wild increase in inaccuracy. This research paper proposes to use Microsoft Azure Cognitive services and Cloud system for implementation of the proposed system [21].
The next phase this research will try to compare this proposed methodology with traditional techniques like HAAR cascade to judge performance of the proposed System, as it is important to have a high accuracy , for a project of this sensitivity.

***Keywords*— Face Recognition, Face Detection, Microsoft AZURE Cognitive Services, Face API, HAAR, Machine Learning, Cloud.**

I. INTRODUCTION

Crime dept. finds it difficult to analyse all public camera video data after a major/minor crime takes place. Finding the past history and movements of the criminal becomes important but manual process is time consuming.Our Paper can help improve this process by reducing search times. By using AI and the power of cloud computing we aim to finish this task faster than any manual process can. Face is a window to human personality, emotions and thoughts. To convey information easily in the field of networking and robotics. Here we develop an application which can find human faces in a video specifically CCTV data, this can be used to find criminals or missing people. The process involves sending individual frames of a video to the Azure cloud service to detect and recognize faces[21]. It takes as input static images of a single user, and finds the face and recognizes it if any. The static images are inputted from a computer program and are stored in database and later processed on cloud[21]. It recognizes a person using Azure cognitive services face API. The formatter will need to create these components, incorporating the applicable criteria that follow [21].

II. LITERATURE SURVEY

This research first surveyed almost twenty research paper which mainly included IEEE papers. This survey helped us helped us gather lots of information the domain of our interest, especially in the field of image processing, detection & recognition. To get an in-depth idea about the topic numerous papers were studied that were written in the field of Face Recognition, Face detection, object detection. Neural networks, image processing and object recognition. This section summarizes some of the papers that influence some of the design and flow of the our proposed methodology & use in selection of algorithm to implement the various module involved in this proposed methodology.

Microsoft Kinect Sensor is used for detecting faces[1]. Kinect has a RGB sensor and a IR sensor. CCTVs have a RGB sensor with IR LEDs for night vision. The device detecting the faces was Kinect with its vast hardware and sensors. These faces were passed on in a cropped manner to the system. Used a Ann to train the system using 4 faces. The Ann was trained on Azure. The processing was done by a PC104+ embedded system. The link between the PC104+ system and the Windows Azure cloud infrastructure was a 3G wireless connection. The system was overall very fast with speeds sometimes around 200ms.

The method used several sub methods like skin detection although he most prominent was face detection [2].The Algorithm used was the Haar Cascade Classifier. HAAR Cascade classifier mainly works by finding patterns with reference to black and white rectangles. This particular

978-1-5386-9353-7/19/$31.00 ©2019 IEEE

paper used a advanced implementation which also involved diagonals. Every Face is modeled as an ellipse. Skin detection is used to speed up haar face detection .This is only detection not recognition.

It is a multistep process where every frame was analyzed [3]. First step was detection using Haar classifier. Second step was detecting face using both Eigen Faces and Gabor algorithm, third step was decision making and selecting. The accuracy was around 50 percent because of changes in illumination pose…etc.

The authors used method which specifically used Microsoft Azure to compute data [4]. This project main aim was to find objects in the environment for robot to navigate. The paper used the SIFT and SURF Algorithm as its main object detection strategy .SIFT finds objects by identifying blobs .Another technique is to try a direct match .The user interface was a windows from C# application.

Authors used the method which used a technique to determine the consistent background [5]. This paper made use of 2 cameras to prevent any disturbance from lighting. The one particular method they used was background subtraction. The specific algorithm was the MOG background subtraction. Long term and shorter analysis is done in case someone come back to pick up his luggage. Algorithm used was the MOG background subtraction. Long term and shorter analysis is done in case someone come back to pick up his luggage.

With increasing terrorist activities there was augmenting demand for video surveillance. Mostly images are generally classified based on the value of simple features [6]. It is always better to use features rather than using pixels as feature based systems always operates much quicker than pixel based systems. In this approach the algorithm consists of 3 intermediate steps :

A) By using an intermediate representation for the image, Integral rectangle features can be computed very quickly.
B) Adaboost technique can be used for the construction of classifiers that helps us to separate desired features from the collection of vast no of features. It utilizes a set of positive and negative images to train.
C) Cascading of different classifiers

A full advantage of Pattern Recognition and Image Processing model can be taken with the help of Open Computer Vision (OpenCV). To detect human face and achieve fast face detection of the video, mosaic gray rules were adopted. Three types of main algorithm based on mosaic model are: Gray Rules: The trisection image model according to the organ segmentation of face is established. Integral Image: Rectangle feature can be computed very quickly using an intermediate representation for the image.

TABLE I.

Reference	Publisher	Method used	Strength	Weakness	Tool used	Dataset
[1]	Research Gate 2015	MS KINECT AND AZURE long with ANN	KINECT DOES FACE DATA PROCESSING	Relatively expensive	KINECT Celeron PC	SMALL locally made
[2]	IEEE 2009	HAAR Cascade	Relatively quick	Only detection	PYTHON	Haar cascade
[3]	IEEE 2016	USED HAAR , Gabor , Eigen faces	Quick	Very inaccurate	PYTHON IDE	Haar predefined
[4]	Springer-Verlag Berlin 2011	Used SIFT ,SURF	Very detail	Little slow	PYTHON IDE	Proprietary
[5]	IEEE 2015	MOG background subtraction	Fairly accurate	Little slow and expensive	MATLAB	Airport images
[6]	IEEE 2014	Integral Image, Adaboost, Cascading	High Detection and Accuracy	Relatively expensive	MATLAB, Dell i5 Laptop	NA
[7]	IEEE 2015	Gray Rules, Integral Image	High Accuracy	More Complex	Pentium 4 PC	NA
[8]	(IJIREEICE) 2016	COTS	NA	NA	NA	Longitudinal Database PCSO_LS Database
[9]	UMASS 2016	HAAR Classifier, Adaboost	Robust performance and high speed	Only detection	NA	NA
[10]	IEEE 2016	CNN	Accurate	Takes time to train	Nvidia Tesla K40 GPU	NA

Frequency Histogram: It removes the non-face region from the image and merge the overlapping face region.

Changes in the facial appearance occurs due to natural ageing of human [8] [22]. There are 147,784 operational mugshots of 18,007 repeat criminal offenders in the longitudinal face database. It consists of at least 5 face images of a subject that were collected over at least a 5year time spa. As elapsed time increases between two face images, population mean trends in genuine scores is estimated by using multilevel statistical models. COTS-A performance is better than COTS-B.

Haar classifier is used to detect faces in the frames that are come as input [9]. A face detection method based on AdaBoost was adopted in the paper. Improved AdaBoost algorithm achieves more robust performance. And high speed over conventional AdaBoost based methods. Study of locality preserving projection is essential and analyzed the impact of LP/P over existing face recognition techniques.

Proposed a method using Bilinear CNN [10]. At each location convolution layer outputs of two CNNs of the image are multiplied. Image labels are used to train bilinear CNN model & it requires training network.

Authors proposed a method using deep learning. Automatically collecting and labelling the data from CCTC videos is done in order to construct a dataset [11]. Face were recognized using VGG Face Recognition Algorithm. Haar Classifier is used for detecting faces. The paper reported accuracy of 99.2%.Training is an essential part and in this case required 2.6M images of 2.6 lakh people .

This paper classify object detection into many categories like model based system, image invariance method, example based method , static object detection , moving object detection .Recursive and non recursive algorithm is

TABLE II.

Reference	Publisher	Method used	Strength	Weakness	Tool used	Dataset
[11]	IEEE 2016	Deep Learning	Accurate Dataset	More time for training	Python	Predefined VGG Dataset
[12]	ICECAT 2017	HAAR Cascade, CamShift Algo	Detect Suspicious Human Behaviour	Higher key frame required	MATLAB	NA
[13]	IEEE 2016	Reccursive & Non Reccursive Algo	Detect Suspicious Human Behaviour	Higher key frame required	NA	NA
[14]	IEEE 2015	IAPRA	Does not affect with variation in pose	Slow	NA	IARPA Janus Benchmark
[15]	University of Oulu 2013	Conditional Random Field	Stable	Takes more time	Python IDE	Pretrained CRF
[16]	IEEE 2016	CIE-Luv Color Space,Variance Formula, Euclidean Distance	CIE system classifies color accrding to the HVS	Use of Algo & Size	NA	NA
[17]	ECE Karunya University 2013	Viola Jones Algorithm, HAAR Wavelet, CNN	Overcoming many challenges in preprocessing stages	Does not support multiple face detection	NA	YouTube Celebrities Dataset
[18]	University of Oxford 2015	Deep CNN	Very Accurate	Large dataset required to train	Python	Labeled Faces in the wild dataset (LFW) YouTube Face(YTF)
[19]	ICCSP 2015	Medium Filtering	Very Detailed	Slow	MATLAB	NA
[20]	IEEE 2010	Gaussian Mixture Model	Fairly Accurate	High Cost & Size	NA	PETS 2006 Dataset, I-Lids Dataset, Result of removed images

used for removing background [12].Object can be tracked using point tracking ,kernel tracking, edge detection and color. This paper classify object detection into many categories like model based system, image invariance method, example based method , static object detection , moving object detection [13].Recursive and non recursive algorithm is used for removing background .Object can be tracked using point tracking ,kernel tracking, edge detection and color.

This paper proposes three steps. Steps include face detection, face feature extraction and finally face recognition [14]. Face was detected and background removed, face features like face cuts and angles were formatted and styled, while recognition goes ahead and identifies it 3 detection methods were experimented with namely camshaft algorithm, Haar classifier, vend finding via motion. Camshift and finding via motion were fast but the most accurate and reliable was Haar classifier.

Authors introduces the Benchmark(IARPA Janus) A (IJB-A), face images are manually localized by publicly available media in the wild dataset (500 subjects) [15] [23]. IJB-A protocol focuses on: the ability to search for a person's face in a set of images (search), and (ii) the ability to compare facial imagery from two persons, and verify whether or not they are the same person (compare). All faces have been manually localized and have not been filtered by a commodity face detector is a key distinction between this dataset and previous datasets. In the IJB-A dataset amount of variation in pose, occlusion and illumination is unprecedented.

Based on conditional random field (CRF) which is combined with the saliency measure which introduces a new salient object segmentation method [16] [24]. Statistical framework formulate saliency measure and contrast local features in illumination, color and motion information. Method is efficiently implemented by using statistical frame work [22]. The integral histogram approach and graph cut solvers effectively implements the method. The feature includes lab color values and optical formation which are obtainable in real time.

In the process for detecting the face from the frame that is extracted from the video, author proposed CIE-Luv color space model [17]. The model is capable of separating out skin and non-skin regions and detect skin like reasons. Skin detection is carried out on the basis of image size (height/width>=2/3).Face verification: He says once the skin detection is done, by using variance formula we can distinguish between face and other parts of body detected in image. The extracted face from face detection step is verified in database faces to find person and return best match location along with face recognized. In proposed system the best thing is video is changed to CIE-Luv color model that works on HSV (Human Vision System) thus accuracy is guaranteed. Variance of test face is calculated and variance of each face in codebook is calculated, as per the variance of test face, near valued face from codebook is selected.

For face detection viola jones algorithm is used to reduce the computational complexity.[18]. This algorithm is split into three modules:
1] Integral image: deals with skin color, motion, facial appearance.
2] For face extraction- illumination, noise low regulation are major problem.
3] Problems are overcome by various techniques such as geometric based. Haar wavelet is used for verifying face

This paper shows two contributions that are made:
A) Assembling of a large scale dataset.
B) Use of deep CNN with appropriate training.
Automation and human combined in the loop to represent large dataset. Data purity and time are the main focus [19].

The face benchmarks (LFW and YTF) are used to achieve comparable state of result [25]. CNN feature extractor, a learnable function obtained by composing several linear and nonlinear operators is used. Euclidean distance used for comparing face in data book and nearest valued face from data book is returned.
The Deep face work was extended by the DeepID. Dataset collection included:
1) A list of candidate identity names are bootstrapped and filtered.
2) More images for each identity are collected.
3) Automatic filter is used to improve the purity.
4) Near duplicate removal
5) Final manual filtering.
6) Training: Learning a face classifier, learning a embedding using a triplet loss. The goal is to minimize the average prediction log-loss .

In distinctive video frames, Author's proposed algorithm will track the real time moving objects [20] [25]. The objective is to relate target objects in consecutive video frame for tracking purpose. Moving object in different frames of video is tracked in the real time using color feature and motion. In image processing median filtering is used to decrease noise which is a nonlinear operation. Target representation and localization and filtering and information association is vital in visual tracking.

III. PROPOSED METHODOLOGIES

A) DECENT ACCURACY WITH LOW COST

The cost associated with the API meant there was a need to find a cost reducing approach as well. Doing local face detection on hardware could reduce the total number of API calls. The accuracy of local detection had to be tested. It was important to check the hardware penalty imposed by local detection.

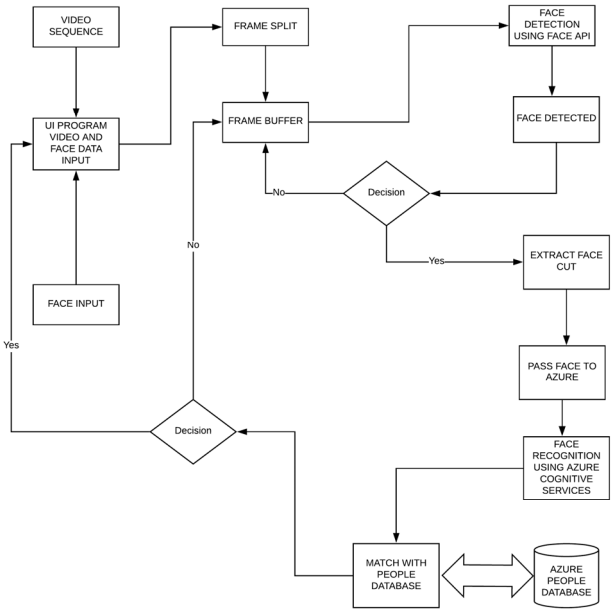

Figure 3.1: Flow chart for face detection using Haar Algorithm

This approach shown in Figure 3.1 is cheaper to implement since it only uses the face recognition service of Azure, but this is taxing on the Computer processor the program it is running .It will be slow considering the laptops or computers this program will be running on will be relatively poor performing.

B) HIGH ACCURACY HIGH COST METHODOLOGY

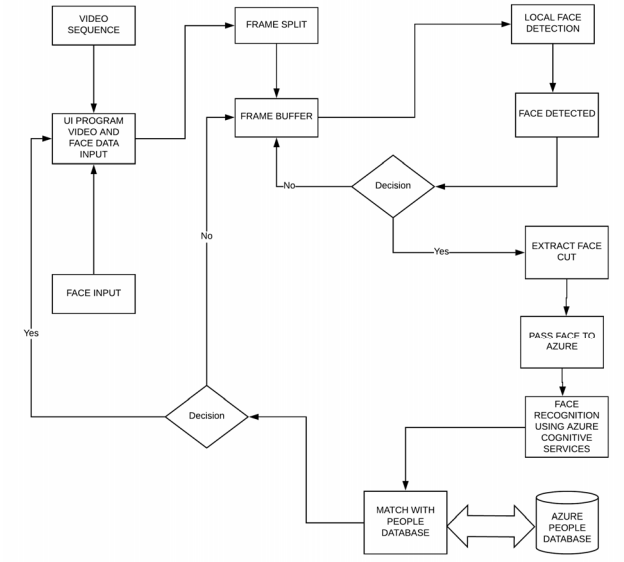

Figure 3.2: Flow chart for face detection using Azure Face API

This approach shown in Figure 3.2 has the ability to provide high accuracy. Azures face API can detect up to 64 faces in a single frame. These than can be extracted and verified with. This also does not stress the computer as it will be running on the cloud. Something very important here is its ability to run on most hardware . On testing this on a system given under a government scheme with relatively poor hardware .it depends mainly on Internet speed . Since most government agencies have leased lines with speeds upwards of 10mbps , this should not be an issue.

IV. CONCLUSION

In this research, we encountered several face recognition algorithms along the way. Some were really accurate while some were really slow. As can be seen in Table I and Table II most implementations are hardware dependent . Our research found that Striking a balance between the Accuracy & Speed is really difficult. Some of the algorithms we came across through means of this research are HAAR , Eigen Faces, Cam shift , CNNs, Viola-Jones Algorithm, Gaussian , Eucledian distance, AdaBoost etc.

The most famous detection algorithm we came across through this research was HAAR. We also came across and learnt about the Microsoft Azure, which bundle of the cloud computing services that are created by Microsoft which can be used for building, testing, as well as deploying, and managing the applications and services using a global network of the data centres managed by Microsoft , which has the various Face API services which is basically cloud-based service which provides algorithms for analysing the human faces found in images and videos. It is fast and accurate and thus our survey and the research helped us choose & use of Microsoft Azure to take care of our face recognition need.

The above proposed approach will be implemented in the next phase of this project which will help us superior verify the performance and high accuracy of this approach with very low latency. Accuracy is of paramount importance for a project of this sensitivity. Also important is the ability to run on any machine regardless of the how powerful the hardware is . These both lead us to choose the Azure face API. The proposed methodology when implemented successfully could definitely be of great help to our Criminal agencies in detecting and finding of criminals, also search for missing people.

REFERENCES

[1] Dobrea, D. M., Maxim, D., & Ceparu, S. (2013, July). A face recognition system based on a Kinect sensor and Windows Azure cloud technology. In *International Symposium on Signals, Circuits and Systems ISSCS2013* (pp. 1-4). IEEE.

[2] Abin, A. A., Fotouhi, M., & Kasaei, S. (2009, October). Real-time multiple face detection and tracking. In *2009 14th International CSI Computer Conference* (pp. 379-384). IEEE.

[3] Tathe, S. V., Narote, A. S., & Narote, S. P. (2016, December). Face detection and recognition in videos. In *2016 IEEE Annual India Conference (INDICON)* (pp. 1-6). IEEE.

[4] Lorencik, D., Ondo, J., Sincak, P., & Wagatsuma, H. (2015). Cloud-Based Image Recognition for Robots. In *Robot Intelligence Technology and Applications 3* (pp. 785-796). Springer, Cham.

[5] Lin, K., Chen, S. C., Chen, C. S., Lin, D. T., & Hung, Y. P. (2015). Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance. *IEEE Transactions on Information Forensics and Security*, *10*(7), 1359-1370.

[6] Chatrath, J., Gupta, P., Ahuja, P., Goel, A., & Arora, S. M. (2014, February). Real time human face detection and tracking. In *2014 international conference on signal processing and integrated networks (SPIN)* (pp. 705-710). IEEE.

[7] Wu, L., Wu, P., & Meng, F. (2010, August). A fast face detection for video sequences. In *2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics* (Vol. 1, pp. 117-120). IEEE.

[8] Best-Rowden, L., & Jain, A. K. (2018). Longitudinal study of automatic face recognition. *IEEE transactions on pattern analysis and machine intelligence*, *40*(1), 148-162.

[9] Chavan, V. D., & Bharate, A. A. (2016). A review paper on face detection and recognition in video. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, *4*, 97-100.

[10] Chowdhury, A. R., Lin, T. Y., Maji, S., & Learned-Miller, E. (2016, March). One-to-many face recognition with bilinear cnns. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-9). IEEE.

[11] Wang, Y., Bao, T., Ding, C., & Zhu, M. (2017, June). Face recognition in real-world surveillance videos with deep learning method. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)* (pp. 239-243). IEEE.

[12] Mishra, P. K., & Saroha, G. P. (2016, March). A study on video surveillance system for object detection and tracking. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 221-226). IEEE.

[13] Goyal, K., Agarwal, K., & Kumar, R. (2017, April). Face detection and tracking: Using OpenCV. In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)* (Vol. 1, pp. 474-478). IEEE.

[14] Rahtu, E., Kannala, J., Salo, M., & Heikkilä, J. (2010, September). Segmenting salient objects from images and videos. In *European conference on computer vision* (pp. 366-379). Springer, Berlin, Heidelberg.

[15] Klare, B. F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., ... & Jain, A. K. (2015). Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1931-1939).

[16] Heshmat, M., Abd-Elhafiez, W. M., Girgis, M., & Elaw, S. (2016, December). Face identification system in video. In *2016 11th International Conference on Computer Engineering & Systems (ICCES)* (pp. 147-154). IEEE.

[17] Yadhul, K., Lakshmi, P. S., & Saju, A. (2014, February). Face detection and recognition with video database. In *2014 International Conference on Electronics and Communication Systems (ICECS)* (pp. 1-5). IEEE.

[18] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015, September). Deep face recognition. In *bmvc* (Vol. 1, No. 3, p. 6).

[19] Singh, P., Deepak, B. B. V. L., Sethi, T., & Murthy, M. D. P. (2015, April). Real-time object detection and tracking using color feature and motion. In *2015 International Conference on Communications and Signal Processing (ICCSP)* (pp. 1236-1241). IEEE.

[20] Tian, Y., Feris, R. S., Liu, H., Hampapur, A., & Sun, M. T. (2011). Robust detection of abandoned and removed objects in complex surveillance videos. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *41*(5), 565-576.

[21] https://azure.microsoft.com/en-in/services/cognitive-services/face/

[22] https://biometrics.cse.msu.edu

[23] https://www.cv-foundation.org

[24] https://citeseerx.ist.psu.edu

[25] Deng, W., Chen, B., Fang, Y., & Hu, J. (2017). Deep correlation feature learning for face verification in the wild. *IEEE Signal Processing Letters*, *24*(12), 1877-1881.