Faces Tracking from Multi-Surveillance Camera (FTMSC)

Hasanen S. Abdullah[1,*], Sana A. Jabber[2]

[1]*Department of Computer Science, University of Technology, Baghdad, Iraq*

[2]*College of Management and Economics, University of Al-Muthanna, Al-Muthanna, Iraq*

[*]*Corresponding author:*qhasanen@yahoo.com

Received: 11-9-2017, Accepted: 13-11-2017, Published: 10-12-2017

DOI: 10.52113/2/04.02.2017/23-32

Abstract: The development of a robust and integrated multi-camera surveillance device is an important requirement to ensure public safety and security. Being able to re-identify and track one or more targets in different scenes with surveillance cameras. That remains an important and difficult problem due to clogging, significant change of views, and lighting across cameras. In this paper, traditional surveillance systems developed and supported by intelligent techniques. That system have ability to performance the parallel processing of all cameras to track peoples in the different scenes (places). In addition, show information about authorized people appearing on surveillance cameras and issue a warning whistle to alert security men if any unauthorized person appears. We used Viola and Jones approach to detected face, and then classifying the target face as one of the authorized faces or not by using Local Binary Patterns (LBP).

Keywords: Face Recognition, Face detection, Viola Jones algorithm, Local Binary Patterns (LBP), video analysis.

1. Introduction

Newly, the spread availability of cameras and rising need for public and private integrity have shifted the interest of researchers about face recognition in video surveillance from multi-camera, this work began 20 years ago [1]. In computer vision and video surveillance, discrimination the face from multi-video has various application. The pose, differences in lighting and unexpected changes in object movement that can vary significantly from frame to frame that is essential challenge of image face detection in videos [2, 3]. On the other hand, distinguishing the faces from the video can give extra precision for the recognition although of the added difficulties in computer vision and video surveillance where a weak recognized in one frames, could be inserting over runs long of video, leading to the power of discrimination [4, 5].

In some cases, like video surveillance, there are giving reasons for capturing and sharing obtain multimedia data to authorize personnel, consequent to security reasons. In most scenarios, it is detect the movement and data on persons in these activities is a demand [4].

In simple words, the Faces Tracking from multi-surveillance camera (FTMSC) proposed system is analysis and distinguish the accurate face in frame from the surveillance multi-video even though there are pose variation. The essential objective of this research is to track object(s) based on detect and recognize the face frame from the multi-video. After extracting the face frame is comparing with the Database, if known, restore the information related to the person in the frame from the database and display it. Otherwise, issuing a warning whistle to alert the security men.

2. Literature Survey

Recently expansion in surveillance-system technology, led to increasing interest in the building of intelligent surveillance systems aimed at public facility management. An intelligent surveillance system is consists of several different proceedings like the object detection, classification, identification, tracking, and information management. The

following are some of the previous work related to this research:
In 2014, the researchers have proposed an intelligent system of attendance management based on the Viola–Jones approach and discrimination partial facial for both controlled environment and uncontrolled. In spite of the reality that the offered system has arrived a high of accuracy within the controlled environment, its achievement within an uncontrolled environment has been actually lower (60%). It has been observing that the rate of facial recognition is different from frame to another frame [7]. In 2016, researchers have suggested a system based on real time face recognition, which is dependable, safe and rapid in conditional lighting conditions [8]. In 2016, the authors have proposed an innovative approach that can super resolve the image with the use of deep learning convolutional network followed by the Hidden Markov Model and Singular Value Decomposition based facial recognition. The results derived from the study demonstrated that the rate of identification is vastly improving after applying the super resolution [9]. In 2017, the authors have proposed a novel investigation algorithm from the facial which beginning with choosing frames, which have a feature-rich from a video sequence using discrete wavelet transform and entropy computation. The outcomes are evincing the algorithm produce the investigation accuracy of over 97% at 1% false positive rate whereas on the YouTube Faces database, over 95% investigation accuracy are spotted at same error rate [6].

3. Architecture of the Proposed System

Conception of the proposed system is be based on the protection of significant places by employ a multi- Security camera put in a central area leading to be protecting the location, the multi security camera is attached to computer system and captures online videos flows and passes it to the proposed system, will be titled to the proposed system "Faces Tracking from multi-surveillance camera (FTMSC)" up to the end of this research. The system architecture is as shown in Figure (1). A FTMSC proposed system, which is using biometrics, face(s) in our case, in general consist of frames capture, Database Design, Pre-processing, Face detection, Feature extraction and classification stages and Post-processing stage. In this paper, a description

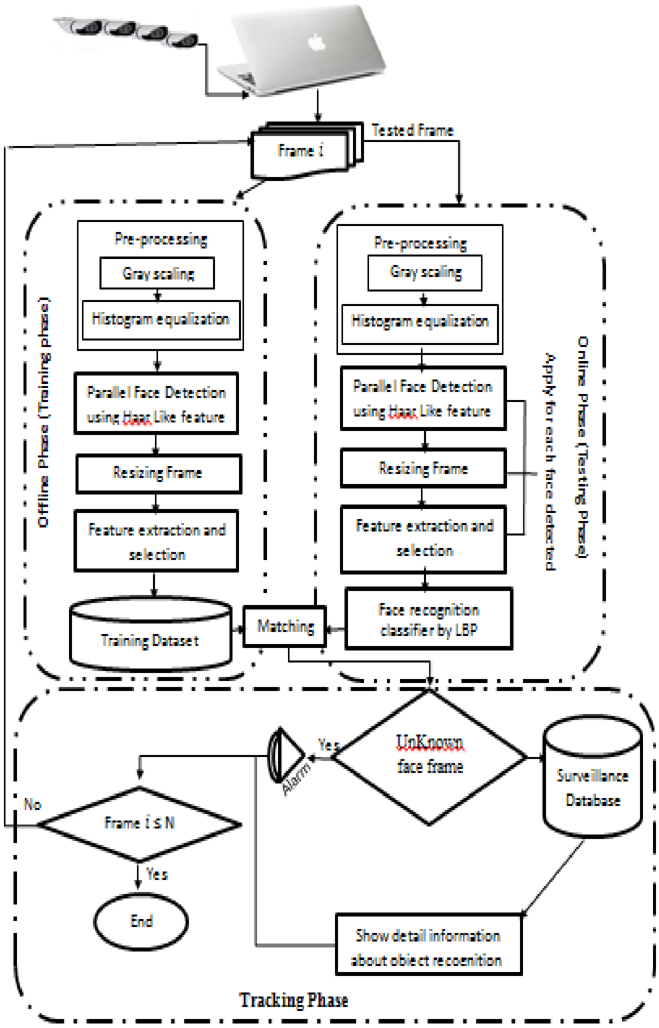

Fig.(1): The proposed FTMSC system architecture.

itemized of the several stages in the proposed model.

4. Capture Frames Stage

4.1. Videos Selecting

Selecting videos module, it utilized to choose an mp4 file from the individual path. To choose the mp4 file by locates IP cameras, and then read the mp4 file.

4.2. Splitting Frame

Splitting Frame module, it used to partition the selection videos file into number of frames.

A video file consists of frames. The frames are the fundamental entity of a video file. Therefore, the frames are splitted and then these frames are stored with an extension .jpg

4.3. Preprocessing Stage

This step includes, convert color frame (RGB) to gray-scale frame and normalized frame by histogram equalization. It considered a Histogram Equalization is the widely common Normalization technique. This enhance the contrast of the frame as it dilates the domain of the intensities in a frame so making it more evident.

4.4. Dataset Design Stage

This Dataset (for a person's face) is design phase that consists of recording each individual's video and extracting the bio-metric feature, in our case it is face, then stored in the database figure (2) show a set of samples face in training dataset. In our study, we took photographs extracting the biometric characteristic, faces of the individuals, then stored in the dataset. We took photographs of individuals at diverse angles, expressions, distances and in different illumination conditions.

Surveillance cameras were install in the room at a height of 2.25 m and placing as shown in figure (3). After the capturing frames procedure, we had to extract only the individuals' faces from the gathered images by the Ada Boost machine-learning algorithm is using to create knowledge base. A database of 35 people with 100 images were collecting for each study. Figure (4) displays the capture of frames from different distances of the camera. Then using the Oracle SQL Plus database to create information database that contains details of information about individuals trained in the face database.

Fig.(2): Set of samples face in training dataset.

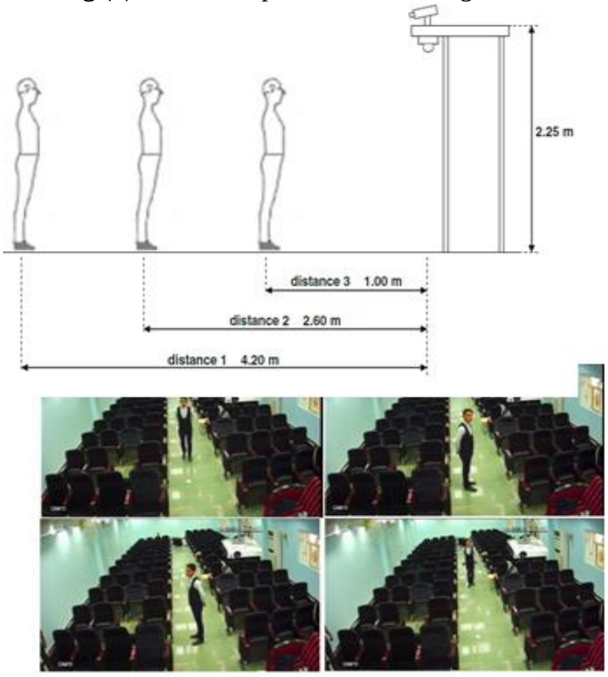

Fig.(3): Capture image in different distance

Fig.(4): Processing of each training framework for the system to authorized Persons. (a) Original frame. (b) Enhancement grayscale frame. (c) Sample of face authorized person after cut from frames.

4.5. The Face Detection Stage:

The efficiency of face recognition systems increases by choosing the suitable and efficient algorithm for detecting and identifying faces in frames, several algorithms are submitted for detection face such as Feature Invariant methods, Face geometry based methods, Machine- learning based methods. Viola and Jones suggested a framework, which grant a high detection estimate, as well fast and robust in real time application [10,11]. For this reason, we chose Viola- Jones framework for face detection, which use of Integral Image, used the most commonly algorithms for feature extraction are Haar-like features and for the classifier process is used an Ada Boost learning algorithm [12].

Papageorgiou, originally used integral image. It grants extremely high speed for calculation rectangular Haar-like features in face detectors. The integral image in (n,m) site is the sum of the pixels that are on top of and left of (n,m):

$$\text{Integral}(n.m) = \sum_{n' \leq n.m \leq m'} original(n'.m') \qquad (1)$$

where the integral image is symbolize as integral (n,m) and the capture image is symbolize as original(n,m) [13].

Haar-like features, Viola and Jones used Haar-like features are the filters to classifies images, which are used features instead of pixels' value directly would increase the speed of the system. When calculate the integral image, will be calculated any one of Haar-like features at any site or scale instantly. Figure (5) offer diverse types of Haar-like features. The sum of pixel values in white rectangles subtracts from the sum of pixel values in gray rectangles. When utilize the detectors with resolution 24x24, the total resulting rectangular features holds approx. 180,000 features [14, 15].

The Ada Boost algorithm is a machine-learning algorithm, which can shape a strong classifier by pick out a small number of the weak classifiers, each one of which is earmark for accurately with one Haar-like feature.

$$h(n) = \begin{cases} 1 \ if \ p_j f_j\,(n) < p_j \theta_j \\ 0 \ otherwise \end{cases} \qquad (2)$$

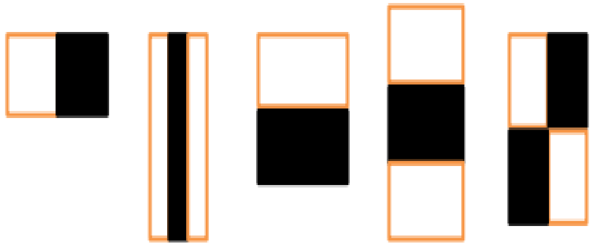

Fig. (5): Examples, rectangular properties shown relative to the enclosing window of detection

where θj is the best or most favorable threshold, fj stands for the used feature, pj for the polarity, which determined classification face as a positive or non-face as a negative. So that the quantum of wrong classified models decreases. Variable n indicates to sub-window of a 24×24 pixel from image. As illustrate before, there are 180,000 feature values, in spite of the fact that need to be choses only a few features. For this reason, the Ada Boost algorithm is using to select that only the best features [16].

Cascaded classifiers: Viola and Jones created a stream architecture of classifiers from a chain of strong classifiers for to speedily rejected unfavorable area of the image. Strong classifiers are arranging in a series at ascending order of complexity [12]. By doing this, there are a large number of areas that are unlikely to contain faces being destroyed by the premier classifiers with few attempt while additional computations are consumed on candidate region by later, more advanced classifiers. This manner increases detection effectiveness and significantly reduces the time of calculation [17].

5. Facial Recognition based LBP

Analysis facial image based on LBP operator is the subject of active research in computer vision in recent years and has effective and widespread applications in many fields, for example the interaction between computer and human, surveillance and security, interactive computer games and identification of identity and etc.[11].

Image pixels with LBP are labeling as a decimal numbers, titled LBP codes that encode the local structure around each pixel. As shown in Figure (6). The values of all adjacent pixels (eight neighbors in a 3X3) is subtracted from center pixels, and Positive and zero values are encoded with 1 and negative values with 0;

associating all these binary codes are clockwise from the top left to get the binary number and use the decimal value corresponding for these binary number for labeling. Referred to Binary numbers as LBP codes or Local Binary Patterns [18].

LBP Extracted image features via segmenting an image into multiple small non-overlapping Z_1, Z_2,.........Z_m (figure 7 shows this) that can be extracted the features for this regions. Those features are madding up of binary patterns representing the circumference of each pixel in the regions. The extracted features from the regions are combined into one feature histogram that which building

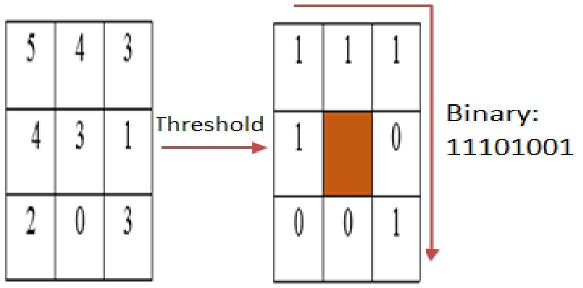

Fig. (6): The original LBP operator

database image, must be computed the difference between the feature vectors via the contrast measurements for histograms.

$$x^2(A,B) = \sum_{j=1}^{L^2} W_j\left(\sum_{i=1}^{p(p-1)+3} \frac{(A_{i,j}-B_{i,j})^2}{A_{i,j}-B_{i,j}} \right) \qquad (5)$$

where $A_{i,j}$, $B_{i,j}$ In above formulas are the sizes of position i from area j (number of occurrence of pattern L(i) in area j), (W_j) weight, the similarity between the two images is calculated by histograms. The lower the x^2 value (It is also known as the "distance" between the two images), It increased the degree of similarity [19]. The algorithm (1) illustrates the overall scheme of the system. Cam1, Cam2, Cam3 and Cam4 represent surveillance cameras, Count, Count1, Count2 and Count3 Count Frames per camera. It is known that the video is composed of a large number of frames and the proposed system deals with more than one surveillance camera so the process of resize of each frame are reduce the complexity of the processing. Confidence represents the similarity between the detected face in the frame and the faces stored in the dataset. To display the frames representation of image. Formally, a pixel from region Z given at (M_c, N_c), the LBP can be explicated in decimal form as:

$$LBP_{X.Y}(m_c.n_c) = \sum_{X=0}^{X-1} s(i_X - i_c)2^X \qquad (3)$$

i_c and i_X are gray-level values respectively of the central pixel and X circumference pixels in the circle neighborhood with a radius Y, and function s(m) is known as:

$$s(m) = \begin{cases} 0 & if\ m < 0 \\ 1 & if\ m \geq 0 \end{cases} \qquad (4)$$

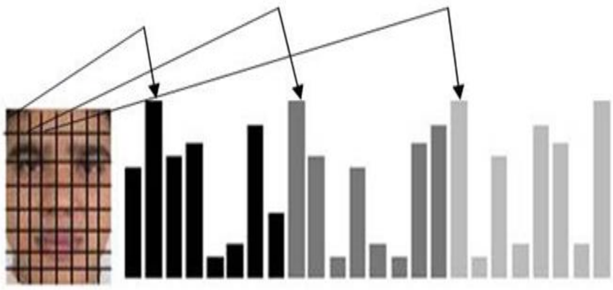

Fig. (7): Facial image divided into 64 area, with very regional hologram.

For recognition facial image LBP comparing between two facial images, input image and after processing on the surveillance screen, the size of frames are resize again and the four frames are distributing on the surveillance screen. Each frame represents a surveillance camera.

5.1. Algorithm 1: General algorithm

Input: video from four-surveillance camera

Output: recognized person authorized or not, display information for all authorized person, issue a warning whistle, frames in my video folder.

Initialize: Count =0, count1 =0, count2 =0, count3 =0

Begin

Video Capture Cam1,Video Capture Cam2,Video Capture Cam3,Video Capture Cam4

If all cameras are not open

print “exit”, end

else while false

spilt any videos to frames

If frame not empty

Begin

3.2.1 Resize frame

3.2.2 Count =count+1
3.2.3 Convert frame to gray scale and equalize (Repeat this processing for each surveillance cameras)
3.2.4 Detection face(s) in frame by using Viola-Jones approach (Repeat this processing for each surveillance cameras
3.2.5 For I=0, I ≤ face number, I + 1
Begin
3.2.5.1 Recognized face detection in frame using LBP
3.2.5.2 If confidence ≤ 60
Begin
3.2.5.2.1 Rectangle face[i]
3.2.5.2.2 If label =1
3.2.5.2.2.1 Pname = "Ali" , print Pname , call Oracle Data Source for display information, end
3.2.5.2.3 If label =2
3.2.5.2.3.1 Pname = "Noor" , print Pname , call Oracle Data Source for display information, end
⋮
3.2.5.2.36 If label =35
3.2.5.2.31.1 Pname = "Mohammed" , print Pname , call Oracle Data Source for display information, end
3.2.5.2.32 else print unknown, alarm
3.2.5.3 end
3.2.6 end
Resize frames to display in screen monetary
3.2.8 Resize (frame1.width/2, frame1.height/2)
3.2.9 Resize (frame2.width/2, frame2.height/2)
3.2.10 Resize (frame3.width/2, frame3.height/2)
3.2.11 Resize (frame4.width/2, frame4.height/2)
3.2.12 frame1 copy to window rectangle (0, 0, 640, 360)
3.2.13 frame2 copy to window rectangle (640, 0, 640, 360)
3.2.14 frame3 copy to window rectangle (0, 360, 640, 360)
3.2.15 frame3 copy to window rectangle (640, 360, 640, 360)
3.3 end
End

5.2. Facial Tracking Phase

Facial tracking (FT) approaches permit following the movement of every one of the Individuals and to re-group facial areas of a same individual (without having knowledge of their identity). Assuming that video streams are recorded with the use of more than a single video camera. The procedure of segmenting performs an isolation of the face areas of interest from successive frames, and distinguishing properties are obtained for representing faces for tracking and classification. A new track is usually created when an emergent face is acquired far from others, and is identified over consecutive frames with the use of the state of the face area that is being tracked Classification features extracted from each facial regions of interest using Local Binary Patterns, the tracking module comes after moving or expressing distinctive faces over the video frames, while the classification module matches facial regions of interest acquired in video to the system's facial models (dataset), the result identifies as a group of facial areas that have the same ID or which the track has high tracking quality QT. As the tracker follows the face area over the scene, the system of segmentation records high quality facial regions of interest for a number of the frames, permitting to result in a trajectory (a trajectory T is identified through consecutive frames). It is important to take under consideration that the module of segmentation doesn't extract a face area from each frame. The diverse group of facial regions of interest belongs to the same person is identified by the tracker. When the quality of tracking QT falls under a (manually) predefined general quality threshold QT> Threshold (confidence), its trajectory is dropped. Finally, the decision fusion combines track IDs and classification scores for the sake of predicting it wanted person appear in front of the camera.

5.3. The Evaluation Proposed System

The main objective of this evaluation process is to verify our hypotheses: Could obtain a good accuracy to automatic recognation and tracking object(s) in differente scenes by using multi video frome surveillance camera?

These FTMSC systems can detect the frontal faces. If needed a better accuracy, should create our own specialised model (Environment). For taking the frames that used in this evaluation, placed the IP camera in the middle of our laboratory and classroom in a position to watch the corridor between the two ends of the room that is considered the point (detection area) that any person must be pass through it, so than it is possible to detect any person entering the room.

The face frame that captured from sequence frame in video of an unknown identity is compared with face images of known individuals from a large dataset. In the figure (8) it can be seen that the input facial images from frames that used as a input to face recognition phase are given below:

Additionally, figure (9) shows the facial images that are stored in the dataset which compared with the input facial images. If the input face images are found or the more similarities face images are matched in the dataset, the face image will be successfully recognized. In addition, face recognized will be surrounded by a red-colored box. The person's name appears above the box.

Fig. (8): Different frame facial image.

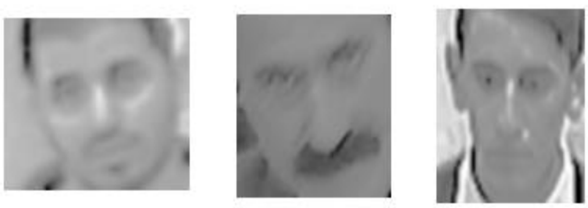

Fig. (9): Facial image from the dataset.

The results of applying the system in a specific environment are displayed in In table (1) and figure (10). 300 frames were captured of video recorded from two-hour surveillance cameras, 60 frame without person and 240 with persons, for detection face in frame to three authorized persons. In this test, detection and recognition rates were calculated separately for each one.

The recognition rates for individuals who had been pre-trained and stored in the dataset are given in table (2) and figure (11) where the rate was calculated from the appearance of persons in four cameras in different scenes with confidence (threshold) equal 65, it is possible to note the variation in the results between one person and another, where observe that the rate of Person ID3 is (95.5) and For Person ID10 is (75.6), face images for thes individual includes many images with severe expressions and noisy images which adversely affect the performance of proposed system. Figure (12) show the ROC curve check the relation between the true positive rate and the false positive rate and figure (13) the PR curve summary the relation between detection precision and the detection rate (recall) in table (2).

Fig. (10): Evaluation proposed system in specific model (environment).

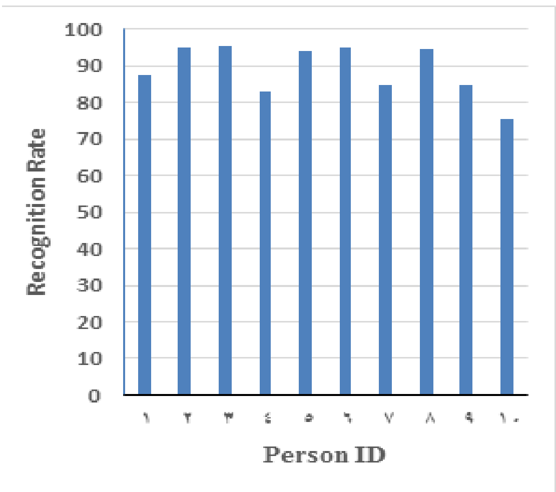

Fig. (11): Recognition rate of persons in proposed dataset.

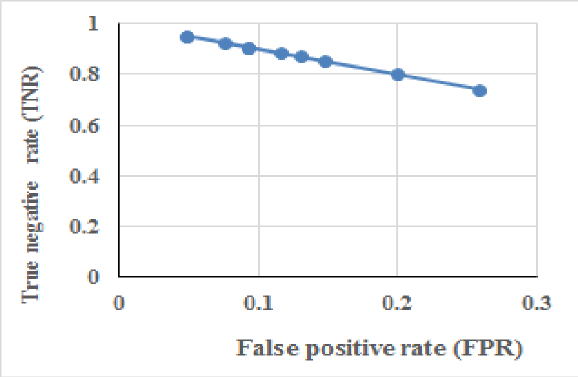

Fig. (12): The ROC curve to proposed system to distinguish ten people within trained dataset.

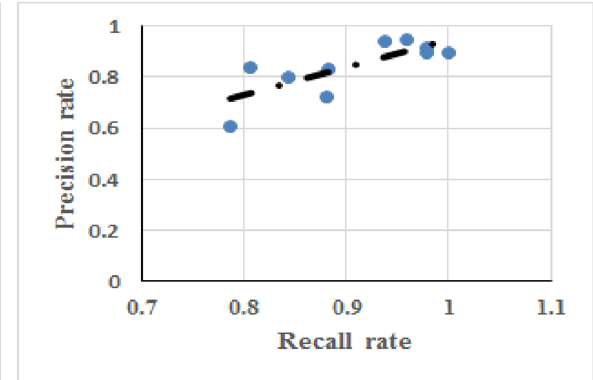

Fig. (13): precision/ recall (PR) curve of face detection methods on proposed system rained dataset.

Fig. (14): Video analysis to track in different scenes.

From figure (14), which is a video analysis to monitor and track a person in four surveillance cameras (video analysis to track the object in different scenes), can answer the question posed at the beginning of section J.

6. Conclusions

This paper proposes a real-time system for surveillance using multi-cameras. The process is splitting into three steps: firstly, the artificial intelligence techniques for training secondly the parallel processing for face detection and thirdly the face recognition to identify particular persons. Some important conclusions are presenting as follows:

The type and resolution of the digital camera plays a fundamental role in the accuracy of detection and recognition.

Lighting plays an important role in detecting and recognition the faces.

Different numbers of training and testing images are using to evaluate the system performance and it shown that increasing the number of training images leads to increases the recognition rate. While reducing the accuracy of recognition rate because of the unrecognized persons within the dataset (strangers).

The system achieved efficient results and high accuracy for such systems that depend on the detection; recognition and tracking face in surveillance multi-cameras are representing by 95% of the detection faces, 89% of recognize and 96% of the tracking of strangers that represent a high proportion compared to similar works.

Table (1): Result of evaluated ISTOMV in specific model

Samples	TN	TP	FP	FN	Precision	Recall	Accuracy	Error rate	TNR	FPR
Person1	100	70	2	4	0.972	0.946	0.966	0.034	0.980	0.020
Person2	98	70	2	2	0.972	0.972	0.977	0.023	0.980	0.020
Person3	100	55	15	0	0.786	1.000	0.912	0.088	0.870	0.130
Average					0.910	0.973	0.952	0.048	0.943	0.057

Table (2): Evaluate the system's to the detection and recognition for Ten persons

Person ID	Precision	Recall	Accuracy	Error rate	TNR	FPR
1	0.833	0.882	0.875	0.125	0.870	0.130
2	0.918	0.978	0.950	0.050	0.925	0.075
3	0.948	0.958	0.955	0.045	0.952	0.048
4	0.725	0.880	0.830	0.170	0.800	0.200
5	0.898	0.978	0.940	0.060	0.908	0.092
6	0.900	1.000	0.949	0.051	0.907	0.093
7	0.843	0.805	0.849	0.151	0.884	0.116
8	0.947	0.937	0.945	0.055	0.952	0.048
9	0.805	0.843	0.848	0.152	0.852	0.148
10	0.611	0.786	0.756	0.244	0.741	0.259
Average	0.843	0.905	0.890	0.110	0.879	0.843

Reference

[1] Xiaoming, Liu; 2005, "Detecting and counting people in surveillance applications", Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on. IEEE.

[2] Rowley, H.; 2008, "Face tracking and recognition with visual constraints in real-world videos", 2008 IEEE Conference on Computer Vision and Pattern Recognition, 06.

[3] Shahabi, C; Kim, SH; Nocera L; Constantinou G; Lu Y; Cai Y-H; Medioni G; Nevatia R; Banaei-Kashani F.; 2014, "Janus–multi source event detection and collection system for effective surveillance of criminal activity", J Inf Process Syst 10(1): 1– 22.

[4] Norris C.; and Moran J.; 2016, "Surveillance, closed circuit television and social control", Book, Routledge.

[5] Yan, Z.; Zheng, X.; and Dai, J.; 2017, "The Big Data Analysis on the Camera-based Face Image in Surveillance Cameras", Intelligent Automation & Soft Computing, 1-9.

[6] Goswami, G.; Mayank V.; and Richa S.; 2017, "Face Verification via Learned Representation on Feature-Rich Video Frames", IEEE Transactions on Information Forensics and Security.

[7] Torre, M. De-la.; Granger, E.; Paulo, V.W. R.; Sabourin, R.; and Gorodnichy, D. O.; 2014, "Partially-supervised learning from facial trajectories for face recognition in video Surveillance", in: Journal of Information Fusion.

[8] Xu, Z.; Mei, L.; Hu, C.; and Liu., Y.; 2016, "The big data analytics and applications of the surveillance system using video structured description technology", Cluster Computing 19, no. 3, 1283-1292.

[9] Rasti, P.; Uiboupin, T.; Escalera, S.; and Anbarjafari, G.; 2016, "Convolutional Neural Network Super Resolution for Face Recognition in Surveillance Monitoring", In International Conference on Articulated Motion and Deformable Objects, pp. 175-184. Springer International Publishing.

[10] Hussain, M.; and Kharat, G.; 2017, "Person Detection and Tracking Using Sparse Matrix Measurement for Visual Surveillance", In Proceedings of the International Conference on Data Engineering and Communication Technology, pp. 281-293. Springer Singapore.

[11] Fathi, A.; Abdali-Mohammadi, F.; 2015, "Improving Face Recognition Systems Security Using Local Binary Pattern", in: Journal of Computing and Security, Vol.2, No.1, pp.55– 62.

[12] Sundaram, M.; and Mani, A.; 2016, "Face Recognition: Demystification of Multifarious Aspect in Evaluation Metrics", In Face Recognition-Semisupervised Classification, Subspace Projection and Evaluation Methods. InTech.

[13] Qiang, L. I. U.; Zhang, W.; Hongliang, L. I.; & Ngan, K. N.; 2016, "Hybrid human detection and recognition in surveillanc", Neurocomputing, 194, 10-23.

[14] García, C. G.; Meana-L., D.; Pelayo B.C.; Bustelo, G.; Lovelle, J. M. C.; Fernandez, N. G.; 2017, "Midgar: Detection of people through computer vision in the Internet of Things scenarios to improve the security in Smart Cities, Smart Towns, and Smart Home", Future Generation Computer System.

[15] Niyomugabo, C.; Choi, H. R.; and Kim, T. Y.; 2016, "A Modified Adaboost Algorithm to Reduce False Positives in Face Detection", Mathematical Problems in Engineering.

[16] Hazim, B. N.; Al-Dabbagh, S. S. M.; Matti, W. E,; and Naser, M. A. S.; 2016, "Face Detection and Recognition Using Viola-Jones with PCA-LDA and Square Euclidean Distance", International Journal of Advanced Computer Science and Applications (IJACSA) 7, no. 5.

[17] Viola, P.; Jones, M.; 2001, "Rabid Object Detection using a Boosted Cascade of Simple Feature", in: Computer Vision and Pattern Recognition.

[18] Xiang, G.; Qiuyu, Z.; Hui, W.; 2016, "Face recognition based on LBPH and regression of Local Binary feature", IEEE.

[19] Reddy, P. K. K.; 2016, "Texture Classification Using Prominent Non Uniform Local Binary Patterns", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Vol.3 Issue: 09.