IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 5, 2018, accepted July 15, 2018, date of publication July 27, 2018, date of current version August 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2856936

Forensic Video Analysis: Passive Tracking System for Automated Person of Interest (POI) Localization

AFTAB KHAN[1], SAKHI REHMAN[1], MUHAMMAD WALEED[1], ASHFAQ KHAN[2], UMAIR KHAN[3], TARIQ KAMAL[1], SHEHRYAR KHAN AFRIDI[1], AND SAFDAR NAWAZ KHAN MARWAT[1]

[1]Department of Computer Systems Engineering, University of Engineering and Technology at Peshawar, Peshawar 25120, Pakistan
[2]Department of Mechanical Engineering, University of Engineering and Technology at Peshawar, Peshawar 25120, Pakistan
[3]Department of Computer Science, COMSATS Institute of Information Technology Attock, Attock 43600, Pakistan

Corresponding author: Aftab Khan (aftab.khan@uetpeshawar.edu.pk)

ABSTRACT Video recorders record the output of each security camera. After an incident, the video footage can be used for evidence by locating a suspect or criminal for a crime. A manual scan of the video footage requires a considerable amount of manpower and time, a luxury which cannot be afforded when tracking down a person of interest (POI). An automated system is proposed in this paper which aims at finding the desired POI through the available volume of video data quickly and accurately. It is visualized to go through all the available videos and detect the POI using facial recognition. Thereafter, it would create a video montage of all the desired frames and incorporate time and location information to produce a path map followed by the POI. The proposed system reduces the human burden, human error and reduces the time taken when searching the POI manually. Validation has been performed on various video data collected by ourselves as well. The results depict that the proposed system is able to correctly identify POI with an accuracy of 86% for video data captured in a constrained environment. Videos captured by a cell phone in an unconstrained environment result in an accuracy of around 80%. Real video tested in our university campus revealed the proposed system is capable of generating tracking information for POI effectively.

INDEX TERMS Video forensics, facial recognition, passive tracking, video summarization, surveillance, security.

I. INTRODUCTION

Public safety is increasingly a major problem in public areas such as educational institutes, airports, malls, railway stations, etc. While major efforts are put in security measures to reduce the crime rate, there is always pressure on the authorities in capturing the culprit once a crime has been committed. Technology has reached a stage where mounted surveillance cameras capture video imagery cheaply. However, having human observers to sit and watch the entire imagery is time-consuming, error-prone and expensive. Nowadays video surveillance systems are installed worldwide in many different sites such as airports, hospitals, banks, railway stations and even at home. The surveillance cameras help a supervisor oversee many different areas from the same room and quickly focus on suspected events taking place in the controlled space. The recorded videos can be exploited to present informative data to the security team who needs to take prompt actions in a critical situation and react in case of unusual events.

As the use of video surveillance system becomes more widespread and the volume of recorded video increases, the need to go through recorded data and extract specific segments and events of interest has become very challenging. After a crime occurs, investigators can go through the video archives to search for information regarding a Person Of Interest (POI). Probing up the video data and looking for the right visual information needs to be performed manually by a human being. It is a laborious task and manually takes a significant amount of time to search for the right video. The search reliability is also questionable as it is subjected to human error. This task is further marred by the presence of a large amount of video data especially in the case when the suspect had arrived at the scene hours before the incident happened.

The benefits of a video surveillance system can significantly increase when it is accompanied by automatic video surveillance and tracking. A practical and efficient surveillance system can be developed by utilizing artificial intelligence and high computing power available today. Utilizing the already available infrastructure effectively is also the main goal of this research work.

An automated system is proposed in this research which is aimed at finding the desired POI by automatically analyzing the available volume of video data. The system is visualized to analyze all the available videos, detect the suspect or POI and make a montage of all the frames from various videos into one single video. The system will reduce the human burden and reduce the time taken in searching the POI manually. The system will generate a single video with all the frames that involves the POI. If camera GPS coordinates are available, it can be used to estimate the path followed by the POI. With this system, the human input will decrease significantly and only a single computerized system can handle multiple cameras and look through volumes of video data quickly.

The automated passive tracking system utilizes facial recognition biometric and machine learning. Section 2 presents the related work in the field of facial recognition, detection and video summarization. Section 3 presents the proposed scheme with a detailed explanation of the different stages and transformations. Experimental results for recognition in high and medium quality videos are given in Section 4 and 5 followed by concluding remarks in Section 6.

II. RELATED TECHNIQUES

The related literature covers the domains of synoptic video, human face detection and tracking, face recognition, human identification in video streaming based on facial classification parameters and motion variation on a Region Of Interest (ROI). The study by Chatrath *et al.* [5] offers a comprehensive empirical analysis of Viola-Jones based face detection scheme. The scheme functions in three steps starting from the development of an integral image, followed by classifier construction for a quick and easy feature selection and in the end cascading these classifiers for reducing the search space and computation time for detection and tracking.

In an effort to recognize faces using more than one camera, some prior work has been done. Xie *et al.* [6] trained a reliability measure and it was used to select the most reliable camera for recognition. Multiple synchronized cameras were used to reduce the pose variance limitations and improved the component based face detector by utilizing Adaboost. They also presented a channel reliability measure to calculate the inherent quality recognition. The improved Adaboost reduces the empirical training error rather than just minimizing the error bound as is accomplished in the plain Adaboost. Harguess *et al.* [7] used a cylinder head model to track and fuse facial recognition results from different cameras. These approaches have been tested on higher resolution videos in a controlled environment in contrast to a low-resolution surveillance video data. Eigen faces is used for stationary face recognition to overcome the effect of transformation errors.

A real-time surveillance system that aims at detecting and tracking humans based on head detection has also been the recent focus of many researchers [5], [6], [8], [9]. One particular study centres on human heads being reconstructed from the fusion of colour histogram and oriented gradients [10]. Though detecting and tracking are two different approaches, Ali and Dailey [11] have merged human detection and tracking into a single algorithm and presented a confirmation-by-classification technique for tracking those detections with tracks. Humans can be identified based on various biometrics including facial, fingerprints and iris [12].

Another technique aims at tracking human beings by generating a motion intensity heat map for a certain period of time [13]. The motion heat map represents hot and cold areas on the basis of motion detection intensities. The hot and cold areas are the zones of the scene where the motion is high and low respectively. Several levels of motion intensity are considered instead of focusing on a single one. The points of interest are then extracted in the selected regions of the scene which are in turn used to estimate motion variations by utilising optical flow techniques for tracking these points.

In forensic video analysis, analysing a complete video is imperative but is time-consuming. A synoptic video system was introduced by Yogameena and Priya [14]. The system generates a relatively shorter length video by preserving only the essential activities from the raw lengthy video. By doing so it aids in reducing the video analysis time through manual inspection. Vural and Akgul [15] presented a surveillance video summarization system, which employed the eye-gaze positions of the surveillance operator and mixes actions from different frames into the same summary frame for more compact videos.

Wang *et al.* [16] showed that useful information for creating video summaries can be extracted by analysing object motion. Rav-Acha *et al.* [17] and Pritch *et al.* [18] also worked on video synopsis in which activities from different time periods are displayed simultaneously by clustering together similar activities. It helps in the simultaneous observation of the objects with similar activities and resulting in outliers being spotted instantly. However, the summaries may be rather confusing when mixing together different activities. Event detection and clustering approach for building both static and dynamic skimming of surveillance video was used in [17] in which events were assumed as the main source of energy change between consecutive frames.

In a preceding study by Kang and Chen [19], Hanjalic and Zhang [20], and Hadi *et al.* [21], two techniques for video summarization are enhanced (i) dynamic video skimming by [19], which itself is still a video but a shorter version, and (ii) static video summary [20], [21], which is a set of images extracted or synthesized from the original video itself. Hamida *et al.* [22] worked on surveillance system and added a pre-analysis stage. The main purpose of this stage is to retrieve compressed information of interest from video footage data. Xu et al. worked on semantic-based model called Video Structural Description (VSD) for depicting and forming the content in videos [23].

Facial Recognition techniques also include Principal Component Analysis (PCA), Discrete Cosine Transform (DCT), Gabor Wavelets method etc. Kaur and Himanshi [24] have utilized PCA method for facial recognition as it provides better recognition rates and less computational complexity. By reducing the dimensionality of the data, a large amount of computation time can be saved which is critical for searching faces in vast image datasets. For each face image, the PCA Eigenvector is almost unique, thus offering fewer chances of false recognition.

DCT like PCA offers a compaction capability which enables us to throw away unnecessary coefficients in the higher range [25]. Sufyanu *et al.* in their work in [26] used anisotropic diffusion illumination normalization technique (AS) in conjunction with DCT for facial recognition. The AS was employed as a pre-processor prior to the application of the DCT as a feature extractor. The AS-DCT technique was able to outperform other benchmark recognition schemes. Multi-resolution wavelet transform combined with DCT and Walsh transform have been used by [27]. The multi-resolution hybrid wavelet transform matrix was generated using Kronecker product of Walsh and DCT transform matrices and then used to extract features from face images with different expressions of subjects' faces. Ajitha *et al.* [28] developed a facial recognition system by using Gabor filters in conjunction with DCT and k-Nearest Neighbour (k-NN) classifier.

Prior to feature extraction, researchers often focus on handling non-uniform illumination problems in the image/video data. A variety of illumination correction techniques have been applied for normalizing illumination to its nominal range [29]–[34].

Section 3 proposes the facial recognition based passive tracking system with a detailed explanation of the different stages and functionalities.

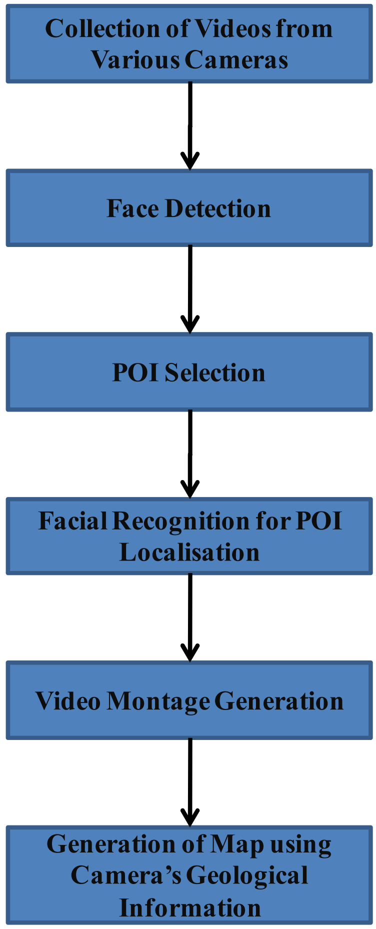

FIGURE 1. Flowchart of the system visualized for POI passive tracking among surveillance videos.

III. PROPOSED PASSIVE TRACKING SYSTEM

The proposed automated system looks for POI in the video data captured by several surveillance cameras. The volume of video data is gathered from different cameras positioned on different locations. Overview of the proposed system visualised for the automatic localisation of POI, generation of synoptic video and path map is given in Figure 1 and 2. Viola-Jones algorithm was employed for face detection while PCA-Eigen Face algorithm [35] has been used for facial recognition. The proposed system includes the following steps:

- All the recorded videos are collected from the available sources e.g. CCTV, PTZ cameras etc.
- Face detection algorithms are applied to the video where the incident took place or started. The Viola-Jones algorithm along with trained classification model for face detection is used to detect faces in the videos.
- The POI is selected from the detected faces.
- Facial recognition is utilised to detect POI from the volume of video data.
- All the frames with a POI hit are indexed and a single video of these frames is generated. The frames in the video are arranged in the chronological order to preserve the sequence of events.
- Utilising the GPS coordinates of each camera and the frames with POI tracked, a map of the path travelled by the POI is approximated using Google Maps [36].

An example scenario illustrating the functionality of the proposed system is presented in Figure 2 for a better understanding of the system's functionality and methodology. Assuming the POI in this case as a suspect who committed crime at Point A and then travels to Point B. The path is covered by four cameras. After the crime, the video recordings from all the cameras are collected with a focus on video data obtained from Camera 1. The recording from Camera 1 is analyzed and the POI is detected. This information is then passed to the proposed system which then analyses all

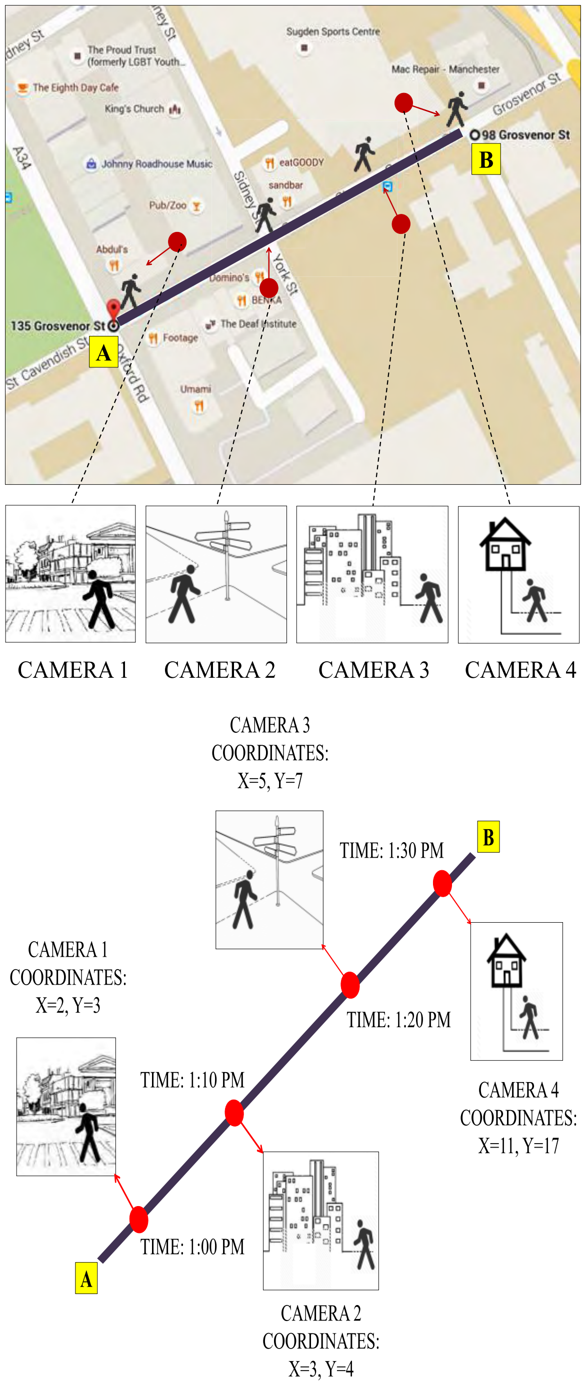

FIGURE 2. (top) Example Scenario with POI travelling from Point A to Point B covered by four CCTV cameras. The example frames with POI are also depicted in the bottom row. (bottom) Video frames with POI arranged in chronological order with the path travelled estimated using the GPS coordinates of the camera.

the videos and searches for the POI among them. Once all the frames with the POI are located, they are combined in chronological order using timestamp information to generate a final video.

If the cameras' geological information is available e.g. the GPS coordinates, the finalized video can be used to generate the path travelled by the POI, as shown in Figure 2. This way the POI can be tracked efficiently and the search time can be reduced a lot.

IV. ADVANTAGES, EXTENSIONS AND OTHER CHARACTERISTICS OF THE PROPOSED SYSTEM

- ***Automatic Alibi Generation*** - Through CCTV footage and utilisation of the proposed scheme, the localisation of a suspect can be easily verified. It can generate a digital alibi from the video footage where the presence of the POI is located.
- ***Active surveillance*** - The same system can be employed for real-time tracking and localisation. The Object Of Interest (OOI) can be a person, car or any object.
- ***Tracking*** - a criminal before and after the crime.
- ***Searching*** - a missing person in a place particularly a building
- ***Movement History*** - Surveillance with multiple cameras utilizing face recognition software and association of the POI with an identity number can lead to the POI's movement being tracked full time and the data recorded against their national identity number.
- ***Scalability*** - The proposed system can be installed one building or facility or it can be extended to a larger region such as a city or entire country.
- ***Complexity*** - Multiple cameras can be used for multiple OOI tracking. This would require an intelligent AI system, good recognition and tracking as well as optic flow image processing systems for prediction of movement.
- ***Privacy*** - Users may not want to be seen or recorded. There is always a trade-off between security and other functional goals.
- ***Usability*** - It can be used by the police, intelligence or security services as well as by a company or organization to track the activities of an employee.
- ***Computational Overhead*** - More hardware and software needs to be installed however pre-existing hardware such as CCTVs can be easily connected by centralized surveillance software.
- ***Resources*** - CCTV or other image recording media have their flaws and limitations. For large scale processing supercomputing will be required.
- ***Recorded Data*** can be useless. Once the user location maps are generated, the video data can be sub-sampled and stored or thrown away.
- ***Reliability*** - The proposed scheme's reliability can be enhanced by integrating other information apart from video footage e.g. GPS on POI's cell phone.
- ***Illegal Immigrants*** can be tracked if national identification database is used. A person not registered in the database would be flagged and tracked till arrested. However, a high amount of resources are required in such case.

The following section demonstrates the experimental setup for the simulations and the results for self-generated as well as benchmark video data.

V. RESULTS AND DISCUSSIONS

The Proposed System was implemented on two different databases:

(i) Self-made videos database for laboratory experiments - the videos were collected from [1], captured in a controlled environment using a high-resolution camera. It also includes videos captured ourselves using a cell phone camera with medium resolution.

(ii) The other benchmark name NRC-IIT Facial Database, were download from [37], for testing. The main goal of creating this database is, to check the computer's capability to identify faces in a situation known to be adequate for humans.

System used for experimentation had the following specifications: Intel(R) Core(TM) i7-3632QM CPU @ 2.20GHz (8 CPUs), ~2.2GHz, 8192MB RAM, 1 GB Virtual Memory with Windows 7 Ultimate 64-bit OS. MATLAB computing language was used as it offers simplicity in image processing algorithm development. It supports a quick and relatively easy coding style for machine learning algorithms. Quantitative measurement of the system's accuracy is defined in the following terms:

- False Negative Rate (FN) - when the face is incorrectly labelled to another person
- True Negative Rate (TN) - when the face is not labelled with any of the available faces
- True Positive Rate (TP) - when the face is labelled with the right person
- False Positive Rate (FP) - when the face is labelled with several persons, one of which is the right person.

The True Positive Rate (TPR) and False Negative Rate (FNR) is defined as TPR=(TP/P), FNR=(TN/P). P and N represent the positive and negative instances respectively. The accuracy of the system is then defined as Accuracy= (TP+TN)/(P+N). Results for videos collected ourselves are presented in the following section.

A. RECOGNITION RESULTS FOR SELF COLLECTED VIDEO DATASET

In order to validate the robustness of the proposed automated system, it was first tested on some self-collected videos including ten videos from [1] and five videos from an iPhone 6 cell phone with sample thumbnails shown in Figure 3(a) and 3(b). Robustness of a system is defined here as the ability of a facial recognition algorithm to produce accurate results over a wide range of data.

A face image is given as input to the proposed system where the face is detected by the Viola-Jones algorithm.

After detecting the dace, the system passes through the entire video dataset and retrieves all images that match the input image. Among the various advantages offered by the proposed system, one benefit that stands out is its independence in producing accurate results even if images carry a facial expression. Figure 3(c) and (d) shows detection results for two videos depicted in Figure 3(a) and (b), one high-quality and one low-quality, respectively. Only a couple of frames are missed in detection for the high-quality video as compared to a higher miss rate in the low-quality video. The high miss rate in the low-quality video is attributed to the low resolution and blurring effect inherent in the capturing medium. Table 1 shows the facial recognition results for the 15 POI in the test videos. The proposed model provides high recognition rate for the ten high-quality videos as compared

TABLE 1. Facial recognition accuracy for high-quality video dataset [1].

POI No.	Ground Truth	TP	FP	TN	FN	TPR %	FPR %	Accuracy %
1	200	120	10	50	20	60.00	5.00	85.00
2	1100	600	65	300	135	54.55	5.91	81.82
3	950	570	23	280	77	60.00	2.43	89.48
4	500	200	14	200	86	40.00	2.80	80.00
5	250	150	31	50	19	60.00	12.40	80.00
6	400	240	25	60	75	60.00	6.25	75.00
7	350	275	21	40	14	78.58	6.00	90.00
8	250	160	5	70	15	64.00	2.00	92.00
9	420	270	17	50	83	64.29	4.05	76.20
10	510	380	15	30	85	74.51	2.95	80.40
11	980	737	8	33	202	75.21	0.82	78.58
12	590	443	4	22	121	75.09	0.68	78.82
13	643	500	2	10	131	77.77	0.32	79.32
14	788	713	3	7	65	90.49	0.39	91.38
15	460	320	6	15	119	69.57	1.31	72.83
Average	559	379	17	81	83	66.94	3.55	82.06

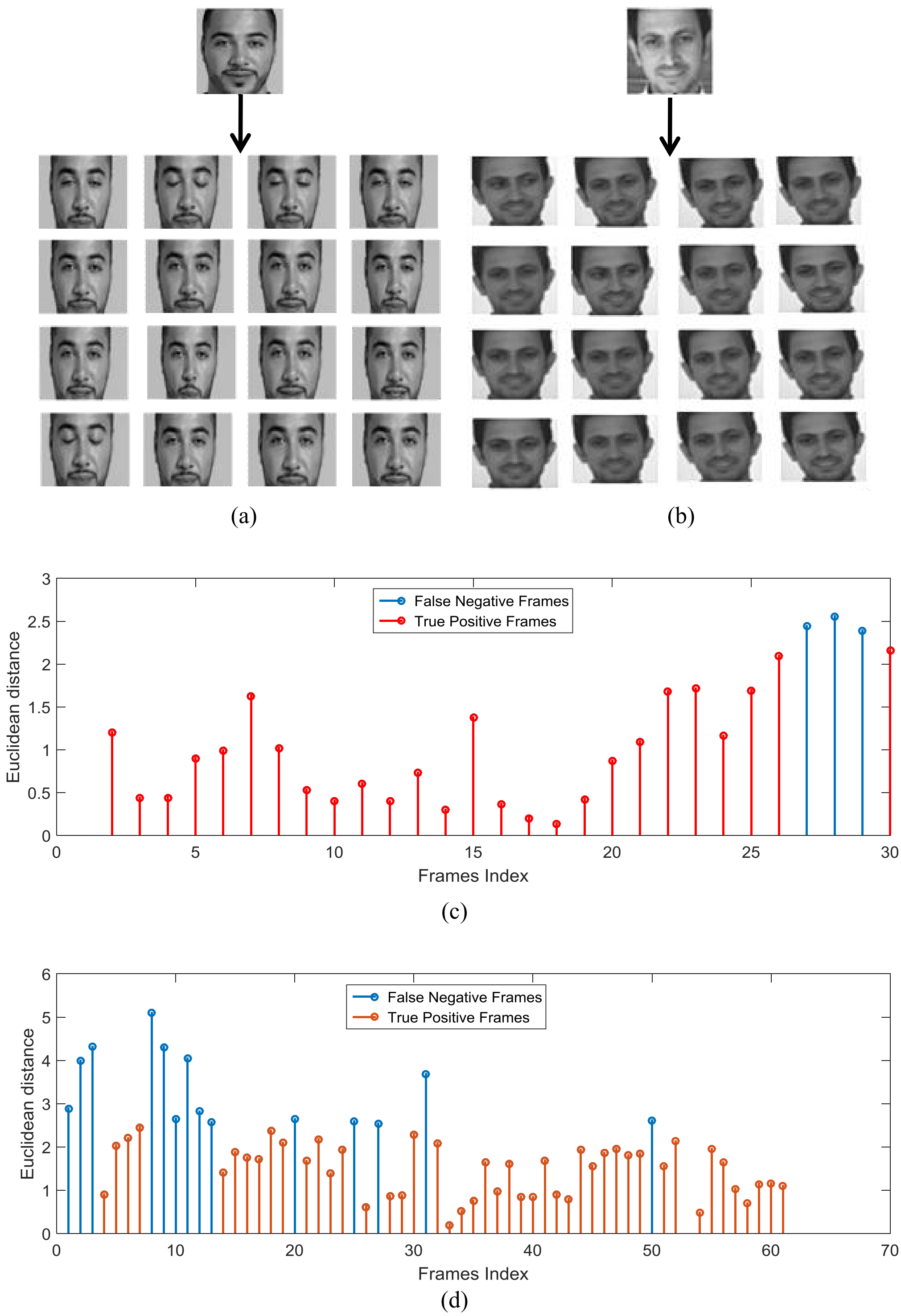

FIGURE 3. **Facial recognition in (a) high-quality video [1] (b) low-quality cell phone video and their respective recognition accuracy. High accuracy is presented for a pristine video as compared to a low accuracy for a cell phone based video.**

to the five low-quality ones. The recognition rate is on average 6.6 percent better than the later one.

B. COMPARISON WITH OTHER BENCHMARK SCHEMES

The proposed system was compared for facial recognition ability against benchmark schemes. In this regard, the proposed system was tested using the NRC-IIT video-based facial dataset, downloadable from [37] and introduced in [2] with thumbnails shown in Figure 4. The dataset contains videos of 12 persons with two videos for each person, one for training (labelled as 0) and one for testing (labelled as 1). These videos were shot in the same background, illumination conditions and camera position.

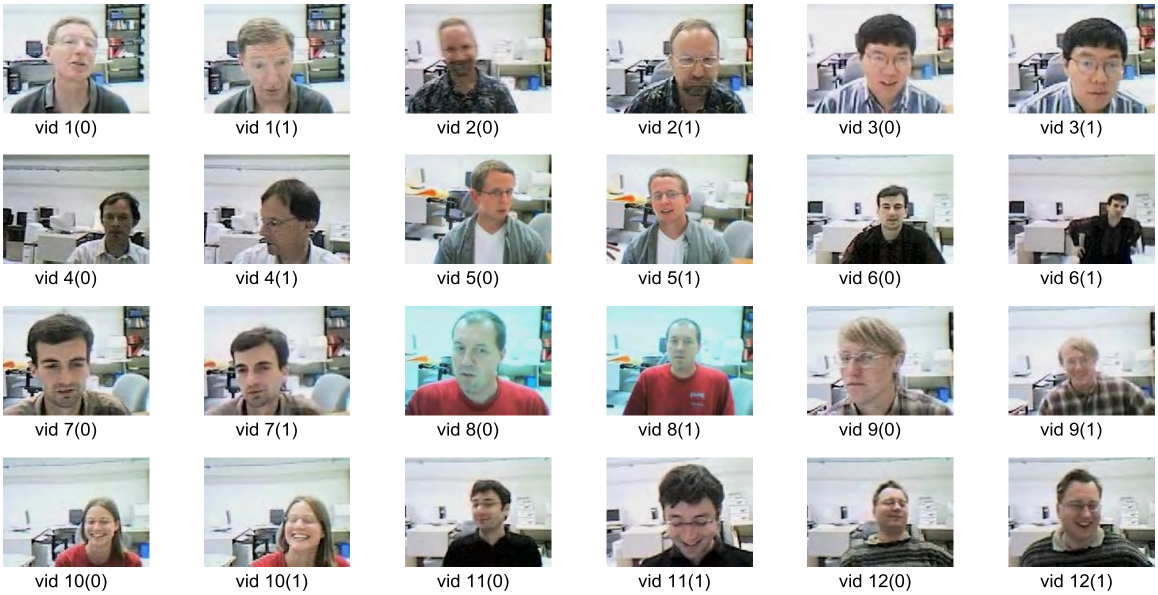

FIGURE 4. NRC-IIT Facial Video Database used for benchmark comparison.

TABLE 2. Comparison of facial recognition accuracy against benchmark schemes and video data.

	Proposed					Gorodnichy [2]				
Video ID	**TP**	**TN**	**FP**	**FN**	**Accuracy %**	**TP**	**TN**	**FP**	**FN**	**Accuracy %**
1	51	0	1	2	94.5	48	0	1	5	88.9
2	165	4	7	9	91.4	160	7	9	9	90.3
3	237	8	15	50	79.0	226	10	18	56	76.1
4	88	5	80	94	34.8	78	7	86	96	31.8
5	24	2	14	82	21.3	20	2	16	84	18.0
6	146	5	20	50	68.3	140	6	22	53	66.1
7	194	22	15	8	90.4	187	25	17	10	88.7
8	243	57	24	75	75.2	235	60	24	80	73.9
9	129	14	40	99	50.7	122	17	42	101	49.3
10	238	9	21	31	82.6	231	12	23	33	81.3
Total	**1515**	**126**	**237**	**500**	**69.0**	**1447**	**146**	**258**	**527**	**67.0**

	Afifi [3]					Struc [4]				
Video ID	**TP**	**TN**	**FP**	**FN**	**Accuracy %**	**TP**	**TN**	**FP**	**FN**	**Accuracy %**
1	46	0	1	7	85.2	44	3	5	2	87.0
2	156	9	9	11	89.2	162	7	7	9	91.4
3	220	10	22	58	74.2	218	10	20	62	73.6
4	90	13	80	84	38.6	80	9	88	90	33.3
5	30	6	16	70	29.5	35	2	16	69	30.3
6	135	6	22	58	63.8	135	4	24	58	62.9
7	180	25	19	15	85.8	170	25	36	8	81.6
8	230	50	29	90	70.2	237	50	24	88	71.9
9	132	17	42	91	52.8	132	22	42	86	54.6
10	235	12	23	29	82.6	221	15	25	38	78.9
Total	**1454**	**148**	**263**	**513**	**67.4**	**1434**	**147**	**287**	**510**	**66.5**

Table 2 shows the recognition accuracy achieved for the NRC-IIT dataset for four schemes including the proposed scheme and three benchmark schemes by Gorodnichy [2], Afifi and Abdelhamed [3], and Štruc and Pavešić [4]. The proposed system has improved accuracy as compared to the benchmark schemes. It achieves an overall accuracy of 69 percent as opposed to an accuracy of 67, 67.4 and 66.5 percent by Gorodnichy, Afifi and Struc respectively. The video resolution is 120 by 160 pixels resulting in a relatively lower recognition accuracy by all the schemes. However, marked improvement is seen in the proposed scheme which is attributed to the fact that we initially up-sampled the video frames to 360 by 480 pixels using bicubic interpolation and a Gaussian low pass filter. This allowed the proposed scheme to detect the faces with less error.

Figure 5 shows the confusion matrices for the proposed and benchmark schemes. It shows that all the schemes produce lower recognition rates for video number four and five. It is due to the pose variations in these two videos which leads to inaccuracy in producing facial features for detection and in turn leading to poor or no recognition.

The proposed system deals with the automated person of interest localization in medium to high-quality video data through utilization of existing surveillance equipment. Its efficiency is however limited by low-quality video data and reduced in unconstrained environments such as varying lighting conditions, non-uniform illumination, rain or fog etc. Currently, it is also challenging to employ the system for

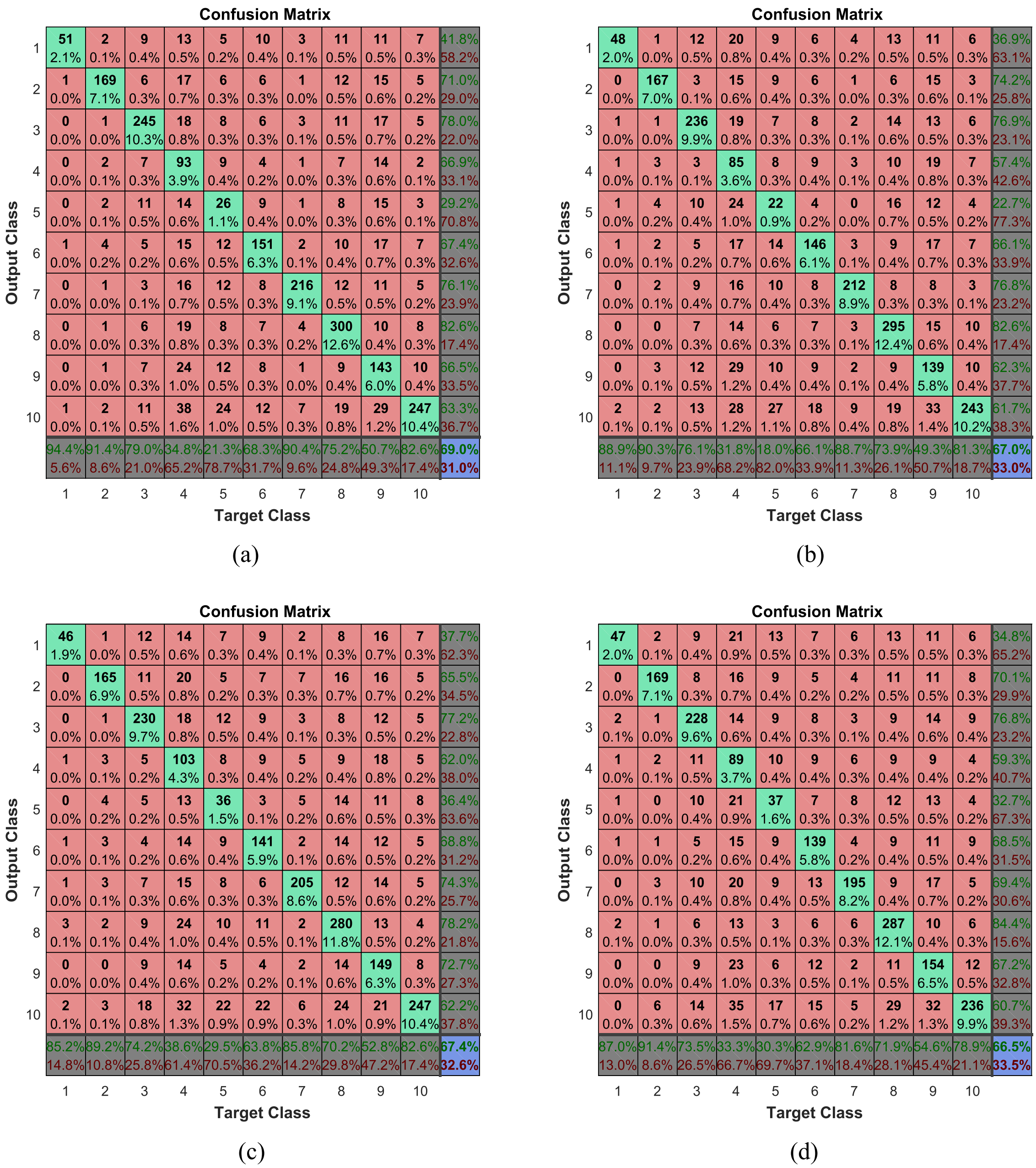

FIGURE 5. **Confusion matrix for the four facial recognition schemes computed for the NRC-IIT Facial Recognition Video Dataset. (a) proposed algorithm (b) Gorodnichy (c) Afifi and (d) Struc.**

facial recognition in a pose-invariant environment as proven by the results.

The system was implemented for real-life testing in our varsity's surveillance system. The implementation details and results are given in the section below.

VI. RESULTS ON CAMPUS SURVEILLANCE SYSTEM

The proposed system was implemented for tracking in our university's security room. It contains 14 cameras including bullet style and high-speed dome cameras. Figure 6 shows the campus map with the cameras locations marked on it in the form of red circles. The arrow depicts the direction in which a camera covers an area. The field of view is limited to 110 degrees and a face can be detected in a maximum range of 30 meters.

The orange path in Figure 6 shows the path designed for the POI and followed during the testing of the proposed system. Camera numbers are also given in the figure. Camera number 4 and 5 are installed on top of the buildings looking down towards the road.

Figure 7 shows the path estimated using the proposed system. Cameras depicted as cyan circles are able to detect,

FIGURE 6. Campus Google map with CCTV cameras locations marked by red circles. The orange track is the path of the POI covered for testing.

FIGURE 7. Campus Google map with CCTV cameras locating the POI (cyan circles) and not localising the POI (yellow circles). Red path with arrows depict the estimated path followed by POI.

recognize and track the POI while cameras marked as yellow circles either do not cover the path of the POI fail to detect him in the video footage of the camera. Camera 4 and 5 is only able to observe the POI from above head height and thus fail to detect him incorrectly. Camera 3 path is blocked by a tree and thus does not detect the POI while camera 12 fails to detect the POI as he is leaving the campus and only a back shot of his body is available. Concluding remarks are presented in the following section.

VII. CONCLUSIONS

An efficient image and video processing analysis system that can automatically localise POI among the recorded videos has been designed and this in combination with the spatial and geological information of the recording medium can generate the path travelled by the POI during a specific time. The proposed system creates a video montage of all the desired frames with POI and incorporates time and location information to generate a map of the path travelled. Comparative experiments indicated that the proposed automated system is able to correctly detect POI with an accuracy of 86 percent for high-quality video data. The system was limited to constrained environment and high quality video footage as low-quality camera data resulted in an accuracy of about 70 percent.

Its performance for facial recognition surpasses other benchmark schemes. Experimentation validates the proposed scheme's facial recognition and passive tracking abilities. The proposed system was also tested on NRC-IIT facial database with the proposed system's recognition at an average rate of 69 percent while the benchmark schemes had a lower accuracy. It was implemented for real-time tracking and results prove the system can easily be integrated with existing security architecture and generates real-time path followed by POI. However, it limited by cameras with either too low-quality or placed at an angle that does not produce a good face image. For future research, further work could be directed towards its improvement in unconstrained environment and computational speed up.

ACKNOWLEDGEMENTS

The authors would like to thank all the reviewers in helping us improve the quality of our research work draft through their valuable advice and efforts.

REFERENCES

[1] Channel. (2016). *100 People O-Face Video*. [Online]. Available: https://www.youtube.com/watch?v=BwDgP6q7lGs

[2] D. O. Gorodnichy, "Video-based framework for face recognition in video," in *Proc. 2nd Can. Conf. Comput. Robot Vis. (CRV)*, May 2005, pp. 330–338.

[3] M. Afifi and A. Abdelhamed. (2017). "AFIF4: Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces." [Online]. Available: https://arxiv.org/abs/1706.04277

[4] V. Štruc and N. Pavešić, "The complete Gabor-Fisher classifier for robust face recognition," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, p. 847680, 2010.

[5] J. Chatrath, P. Gupta, P. Ahuja, A. Goel, and S. M. Arora, "Real time human face detection and tracking," in *Proc. Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2014, pp. 705–710.

[6] B. Xie, V. Ramesh, Y. Zhu, and T. Boult, "On channel reliability measure training for multi-camera face recognition," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Feb. 2007, p. 41.

[7] J. Harguess, C. Hu, and J. K. Aggarwal, "Fusing face recognition from multiple cameras," in *Proc. Workshop Appl. Comput. Vis. (WACV)*, Dec. 2009, pp. 1–7.

[8] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.

[9] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Comput. Vis. Image Understand.*, vol. 80, no. 1, pp. 42–56, 2000.

[10] R. Xu, Y. Guan, and Y. Huang, "Multiple human detection and tracking based on head detection for real-time video surveillance," *Multimedia Tools Appl.*, vol. 74, no. 3, pp. 729–742, 2015.

[11] I. Ali and M. N. Dailey, "Multiple human tracking in high-density crowds," *Image Vis. Comput.*, vol. 30, no. 12, pp. 966–977, 2012.

[12] N. Saparkhojayev and Y. Akhmetov, "Human identification in video streaming based on some facial classification parameters," *Middle-East J. Sci. Res.*, vol. 21, no. 7, pp. 1153–1156, 2014.

[13] N. Ihaddadene and C. Djeraba, "Real-time crowd motion analysis," in *Proc. Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[14] B. Yogameena and K. S. Priya, "Synoptic video based human crowd behavior analysis for forensic video surveillance," in *Proc. 8th Int. Conf. Adv. Pattern Recognit. (ICAPR)*, Jan. 2015, pp. 1–4.

[15] U. Vural and Y. S. Akgul, "Eye-gaze based real-time surveillance video synopsis," *Pattern Recognit. Lett.*, vol. 30, no. 12, pp. 1151–1159, 2009.

[16] Y. Wang, T. Zhang, D. Tretter, and P. Wu, "Real time motion analysis toward semantic understanding of video content," *Proc. SPIE*, vol. 5960, p. 596027, Jul. 2006.

[17] A. Rav-Acha, Y. Pritch, and S. Peleg, "Making a long video short: Dynamic video synopsis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2006, pp. 435–441.

[18] Y. Pritch, A. Rav-Acha, and S. Peleg, "Nonchronological video synopsis and indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1971–1984, Nov. 2008.

[19] H.-W. Kang and X.-Q. Chen, "Space-time video montage," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 1331–1338.

[20] A. Hanjalic and H. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1280–1289, Dec. 1999.

[21] Y. Hadi, F. Essannouni, and R. O. H. Thami, "Video summarization by k-medoid clustering," in *Proc. ACM Symp. Appl. Comput.*, 2006, pp. 1400–1401.

[22] A. B. Hamida, M. Koubaa, H. Nicolas, and C. B. Amar, "Video surveillance system based on a scalable application-oriented architecture," *Multimedia Tools Appl.*, vol. 75, no. 24, pp. 17187–17213, 2016.

[23] Z. Xu, C. Hu, and L. Mei, "Video structured description technology based intelligence analysis of surveillance videos for public security applications," *Multimedia Tools Appl.*, vol. 75, no. 19, pp. 12155–12172, 2016.

[24] R. Kaur and E. Himanshi, "Face recognition using principal component analysis," in *Proc. IEEE Int. Adv. Comput. Conf. (IACC)*, Jun. 2015, pp. 585–589.

[25] Z. Pan and H. Bolouri, "High speed face recognition based on discrete cosine transforms and neural networks," Tech. Rep., 1999.

[26] Z. Sufyanu, F. S. Mohamad, A. A. Yusuf, and M. B. Mamat, "Enhanced face recognition using discrete cosine transform," *Eng. Lett.*, vol. 24, no. 1, pp. 52–61, 2016.

[27] A. Choudhary and R. Vig, "Face recognition using multiresolution wavelet combining discrete cosine transform and Walsh transform," in *Proc. Int. Conf. Biometrics Eng. Appl.*, Hong Kong, 2017, pp. 33–38.

[28] S. Ajitha, A. A. Fathima, V. Vaidehi, M. Hemalatha, and R. Karthigaiveni, "Face recognition system using combined Gabor wavelet and DCT approach," in *Proc. Int. Conf. Recent Trends Inf. Technol.*, Apr. 2014, pp. 1–6.

[29] J.-X. Felix and M. Savvides, "Subspace-based discrete transform encoded local binary patterns representations for robust periocular matching on NIST's face recognition grand challenge," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3490–3505, Aug. 2014.

[30] M. Lee and C. H. Park, "An efficient image normalization method for face recognition under varying illuminations," in *Proc. 1st ACM Int. Conf. Multimedia Inf. Retr.*, Vancouver, BC, Canada, 2008, pp. 128–133.
[31] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," in *Proc. Int. Conf. Biometrics*. Berlin, Germany: Springer, 2009, pp. 209–218.
[32] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
[33] S. Du and R. K. Ward, "Adaptive region-based image enhancement method for robust face recognition under variable illumination conditions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 9, pp. 1165–1175, Sep. 2010.
[34] Z.-R. Lai, D.-Q. Dai, C.-X. Ren, and K.-K. Huang, "Multilayer surface albedo for face recognition with reference images in bad lighting conditions," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4709–4723, Nov. 2014.
[35] A. E. Omer and A. Khurran, "Facial recognition using principal component analysis based dimensionality reduction," in *Proc. Int. Conf. Comput., Control, Netw., Electron. Embedded Syst. Eng. (ICCNEEE)*, Sep. 2015, pp. 434–439.
[36] J. E. Rasmussen and L. Rasmussen. *Google Maps*. [Online]. Available: https://www.google.com/maps
[37] D. O. Gorodnichy. (2005). *NRC-IIT Facial Video Database*. [Online]. Available: http://www.videorecognition.com/db/video/faces/cvglab/

AFTAB KHAN received the B.E. degree in computer system engineering from the College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Islamabad, Pakistan, in 2009. His Ph.D. thesis was on Single-Image Blind Deblurring and Restoration Techniques at The University of Manchester, U.K., in 2014. From 2010 to 2013, he was a Graduate Teaching Assistant with The University of Manchester. He has been an Assistant Professor with the Department of Computer Systems Engineering, University of Engineering and Technology (UET) at Peshawar, Peshawar, Pakistan, since 2013. His image restoration expertise includes restoration of blurred images using blind deblurring techniques based on non-reference image quality measures. His image processing skills include OCT image denoising, medical image analysis, and image fusion. His research interests include digital image restoration, blind image deblurring, medical image processing, and digital image and video compression. He was a recipient of the Faculty Development Scholarship from UET at Peshawar, for his Ph.D. thesis.

SAKHI REHMAN graduated from the University of Engineering and Technology at Taxila, Pakistan, in 2013. He did his Intermediate (Computer Science) from 2007 to 2009 From BISE Kohat. He is currently pursuing the M.Sc. degree in image processing with the University of Engineering and Technology at Peshawar. He is currently with OGDCL as Trainee Computer Engineer. His interests include image processing, networking, and data compression.

MUHAMMAD WALEED received the B.Sc. and M.Sc. degrees from the University of Engineering and Technology (UET) at Peshawar, in 2017 and 2014, respectively, where he is currently pursuing the Ph.D. degree in image processing. He did his Intermediate (Pre-Engineering) from 2008 to 2010 at Government College, Peshawar. His interests include image processing, networking, and data compression. He is currently a Teaching Assistant with UET Peshawar.

ASHFAQ KHAN received the B.E. degree in mechanical engineering and the M.Sc. degree in advance manufacturing from the National University of Sciences and Technology in 2007 and 2008, respectively. He received the Ph.D. degree from The University of Manchester (UoM) , U.K., in 2011. Soon after his Ph.D. studies, he was Research Visitor at UoM, where he was able to continue his research work. He has been an Assistant Professor with the Department of Mechanical Engineering, University of Engineering and Technology (UET) at Peshawar, since 2012. He is the first-ever Ph.D. graduate under UET's development project, UET Peshawar, Jalozai Campus, and has returned before his due time after completing his Ph.D. from UoM, in two years and ten months. He is currently the youngest PhD of Khyber Pakhtunkhwa (KPK). Afterward, he conducted the Ph.D. research on laser based micro/nano manufacturing and characterization at the UoM, U.K., funding by UET Peshawar. His laser processing expertise includes processing with pulsed and continuous wave lasers of different wavelengths. His characterization skills include optical microscopy, scanning electron microscopy (SEM), transmission electron microscopy, X-ray diffraction, atomic force microscopy, and white light Interferometery. His research interests include laser advanced manufacturing, micro/nano fabrication, laser-based net shape manufacturing, CNC machining, rapid prototyping, and micro/nano characterization.

UMAIR KHAN is a graduate student within the Computer Systems Engineering Program at the University of Engineering and Technology (UET), Peshawar Pakistan. He is currently pursuing the M.S. degree in computer system engineering with UET Peshawar, Pakistan. He is currently a Research Associate with the COMSATS Institute of Information Technology Attock. His research concentrates on advanced concepts of image processing and machine learning.

TARIQ KAMAL received the B.Sc. degree in computer system engineering from University of Engineering and Technology (UET) at Peshawar, the M.Sc. degree from The George Washington University, USA, and the Ph.D. degree from Virginia Tech, USA. He is currently an Assistant Professor with the Department of Computer System Engineering, UET Peshawar. His research interests are in the field of software engineering and high performance computing. Specifically, he is interested in using the power of high performance clusters and distributed systems for the applications of large-scale big-data simulations and its impact on socio-technical development. His teaching interests include computer and system programming, advanced data structures, and cloud computing.

SHEHRYAR KHAN AFRIDI received the B.Sc. degree in computer engineering from the COMSATS Institute of Information Technology at Abbottabad in 2007 and the M.Sc. degree in computer system engineering from the University of Engineering and Technology at Peshawar in 2012, where he is currently pursuing the Ph.D. degree with the Computer System Engineering Department. He was with CISNR Electrical Engineering Department, University of Engineering and Technology at Peshawar as a Research Associate. He was also a Research Analyst with the Computer Science Department, Institute of Management Sciences, Peshawar. He has been serving as a Senior Lecturer with the Computer Science Department, Iqra National University, Peshawar, since 2014. His areas of interest are network security, access control, mobile security architecture, and security and privacy in IoT. He is a technical program committee member of journals by Elsevier Science, John Wiley & Sons, and also the IEEE conferences.

SAFDAR NAWAZ KHAN MARWAT received the B.Sc. degree in computer systems engineering from the University of Engineering and Technology (UET) at Peshawar, Pakistan, in 2006, and the M.Sc. degree in communication and information technology from the University of Bremen, Germany, in 2011. He defended his Ph.D. thesis at the University of Bremen in 2014. He was a Lecturer with the Peshawar College of Engineering, Peshawar, Pakistan, from 2006 to 2008. He is currently an Assistant Professor with UET Peshawar. His research focuses on machine-to-machine communications and future mobile networks.