

CCTV Quality Assessment for Forensics Facial Recognition Analysis

Mohamad Firham Efendy Md Senan
Digital Forensics Department
CyberSecurity Malaysia
Selangor, Malaysia
firham@cybersecurity.my

Wafa Mohd Kharudin
Digital Forensics Department
CyberSecurity Malaysia
Selangor, Malaysia
wafa@cybersecurity.my

Siti Norul Huda Sheikh Abdullah
Pattern Recognition Research Group
Center for Artificial Intelligence Technology
Faculty of Technology and Information Science
Universiti Kebangsaan Malaysia
Selangor, Malaysia
mimi@fstm.ukm.my

Nur Afifah Mohd Saupi
Digital Forensics Department
CyberSecurity Malaysia
Selangor, Malaysia
nurafifah@cybersecurity.my

Abstract— Closed-circuit television (CCTV) is used to perform surveillance recordings, and it is one of the most common digital devices that provide digital evidence for the purpose of forensic analysis. In video forensic analysis, the footage with the target subject or object is extracted out from the CCTV recordings for further analysis. However, the quality of these recordings are often poor due to several factors, such as the type of the camera, the configuration, and also the position of the camera. The results of forensic face recognition depend highly on the quality of the CCTV recordings. Poor quality of CCTV recordings would reduce the confidence level of the face recognition result, thus would not make a strong evidence to be presented in a court of law. The objective of this research is to conceptualise a framework for quality assessment in CCTV evidence to be used in forensic face recognition analysis. The method of this research was divided into two phases. Initial phase covered CCTV evidence testing phase where the experiment was done based on different types of CCTV camera with different resolutions, and distances between the subject and the camera. In the second phase, the face of the subjects were compared to the face taken during the enrolment phase. The score obtained from the forensic face recognition system would be based on the camera resolutions, types of camera, distances, and also the changes of ranking score after applying the enhancement process such as Bicubic to the facial images. The results were analyzed for quality assessment towards these parameters. In general, the evaluation of scoring and ranking decreased as the distance increased. The face also could not be detected by the system when they were taken more than 5 meters distance from the camera.

The highest score of 5.95 was obtained by using resolution 1280 x 720 at distance of 3 meters taken by camera model ACTI E62. The Bicubic enhancement method improved the scoring and ranking especially with the camera model that have low resolution modes.

Keywords —CCTV; Video Forensics; Facial Recognition; Quality Assessment; Bicubic Enhancement Method

I. INTRODUCTION

CCTV is a device which main function is to perform recordings mainly for surveillance and security purposes. An obvious use of CCTV is to prevent crimes and malfunctions, but it is also used for other purposes such as to monitor industrial processes and also traffic movement. CCTV records the movement of people and objects in the surroundings within a specific distance and range. In digital forensics investigation, the people or the object will be the focus and become the interested scene. The specific footage will be extracted out from the recordings of the CCTV for further analysis.

The configurations and types of CCTV are often different, as they depend on the way that the CCTV is being setup during the installation at the premise. In face recognition analysis, the quality of the CCTV camera significantly affects the result of the analysis. The specification of the camera and CCTV configuration vary due to its functionality. Some CCTV cameras are made to capture the surroundings and some other cameras are made purposely for biometrics identification such as facial recognition. The output could vary from excellent quality to a very low quality of video recordings.

In common scenarios, analysts would just receive cases from law enforcement agencies without knowing whether the quality of CCTV recordings is acceptable for facial recognition analysis or not. Most CCTV recordings received would show poor quality of the subject face. The resolution could be less than 100x100 pixels, which is very poor quality. The current approach requires this issue to be discussed based on the uniqueness of the case. However, it is important to conduct and construct an evaluation of CCTV quality so that it can provide us information on a set of acceptable range when we receive cases with specific types of CCTV.

The objectives of this paper is to conceptualize a framework for video quality assessment in CCTV evidence to be used in forensic face recognition analysis, and to identify corrective measures on the CCTV evidence issues such as distance, type of CCTV, and set of resolutions.

This paper is arranged as follow. Section two discusses the state-of-art of CCTV forensic facial recognition analysis, while section three describes the methodologies used in this study. Section four provides the results and discussion gained from the trial, and lastly, conclusion and future works are discussed in section five.

II. FACIAL RECOGNITION FOR FORENSICS ANALYSIS

To determine a person's identity, particularly from surveillance imagery, facial recognition technology

is in continuing progress of researching in order to improve the process. Face recognition is defined by [1] as the ability to establish a subject's identity based on facial characteristics. The data are extracted from seized evidences for example, CCTV, laptops, mobile phones, tablets, video sites, and social media. One of the common issues regarding face recognition for forensic analysis is that faces under investigation inside these medias are often partial-face, which is a problem as it will cause low accuracy in the face matching analysis. Recent research on face recognition has been focusing on reducing the impact of nuisance factors such as pose, illumination, and expression variations on face recognition [2]. However, most of the approaches taken are only able to deal with certain kinds of face variation well but there are constraints restricting their application when multiple variations are involved and only one gallery image available per subject. A technique known as Adaptive Principal Component Analysis (APCA) was proposed by [3] for illumination and expression invariant face recognition with only one gallery image.

There are a lot of challenges in the design of quality assessment algorithms for videos and images. This is because the results depend highly with human subjectivity, and since humans perceive very different judgments, perfect correlation is impossible. The main goal of quality assessment is to produce automatic video and image rating that correlate well with Mean Opinion Scores (MOS). [4] referred MOS as the way of assessing the quality of an image by looking at it because human eyes are the ultimate receivers in most image processing environments. However, the MOS method is impractical in many situations as it is said to be too inconvenient, expensive, and slow.

Generally, video and image quality assessment algorithms are grouped into three categories: full-reference (FR), reduced-reference (RR), and no-reference (NR) algorithms [5]. As their name suggests, the group refers to the amount of information available about the original reference signal. Currently, the most commonly used FR objective image and video distortion/quality metrics

are mean squared error (MSE) and peak-signal-to-noise ratio (PSNR). PSNR is used to measure the quality of reconstructed image from the original image. Higher PSNR generally indicates that the output image is of higher quality. They are widely used because simple to calculate, have clear physical meanings, and are mathematically easy to deal with for optimization purposes (Wang et al., 2004). However, they have been criticized as well, for not correlating well with perceived quality measurements [6] and [7]. These traditional video quality metrics are known to disregard the viewing conditions and the characteristics of human visual perception [8].

III. METHODOLOGY

The evaluation of video quality assessment was done on different types of cameras and distances of the subject from the camera. The process flow in Figure 1 explained the evaluation process.

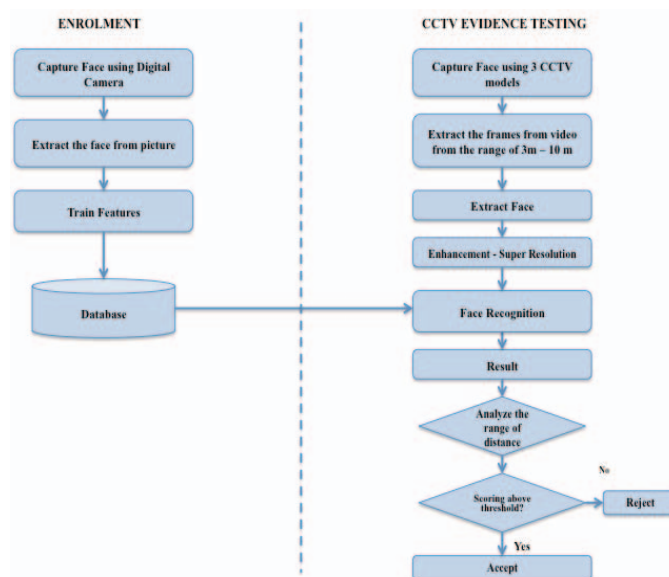


Fig. 1. The evaluation process

There were two phases of the process: Enrolment phase and CCTV evidence testing Phase. In the enrolment phase, the data was collected in a controlled condition. It covers from *Capturing Face using digital camera* until *Database Phase*:

- *Capture Face using Digital Camera* – The subject face was captured using a digital

camera in a controlled environment while facing the camera with dark background. The face was captured in a video format with several positions throughout the recording. However, in this research, we only focused on the frontal face images. Figure 2 shows the example of video frames of the subject facing the digital camera during the enrolment process.



Fig. 2. Example of video frames of the subject facing the digital camera

- *Extract the face from the picture* – Python programming was used in this process to automatically detect and crop the subject face from the video. The face extracted will be used to prepare the database for the face. We chose the best frontal face for the Enrolment phase. Figure 3 is an example of the subject face extracted from the picture.

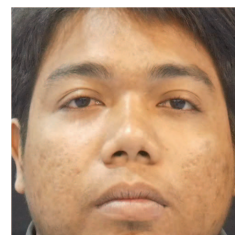


Fig. 3. Example of face extraction

- *Train features* - Once the extraction was completed, the face of subjects will be trained using Forensics Facial Identification System used by CyberSecurity Malaysia (CSM). The system allows us to do the landmark at the face where it needs to be frontal to do manual markup for the landmark process.
- *Database* – For this research we collected the faces of subjects consisted of 4 males

and 6 females. This dataset would be used in the enrolment of the Forensics Facial Identification System.

As for the CCTV Evidence Testing Phase, the data collected in the position of the subject not in the fixed position. This phase covers from *Capture Face using 3 CCTV Models* until *Analysis*.

- *Capture Face using 3 CCTV Models* - For the CCTV evidence testing we used 3 different types of camera with different specifications. The specification would differentiate the quality of the camera. Table I, Table II, and Table III is the description model of camera that we used in this research.

TABLE I. DESCRIPTION OF CAMERA ACTI E62



	
Model	ACTI E62
Maximum Resolution	3MP
Compression	H.264 HP, MJPEG
Max. Resolution	2048 x 1536

TABLE II. DESCRIPTION OF CAMERA RAVEN RPS-AE1

	
Model	Raven RPS-AE1
Video Resolution	1MP
Compression	H.264 HP, MJPEG
Max. Resolution	1290x720

TABLE III. DESCRIPTION OF CAMERA MOBOTIX ALLROUND DUAL M15 DAY NIGHT 90° CAMERA

	
Model	Mobotix AllRound Dual M15 Day Night 90° Camera
Video Resolution	6 MP
Compression	No Compression
Max. Resolution	3072 x 2048

- *Extract the frames from the video from the range from 3m to 10 m* – For this step, we start from the first camera with highest resolution of the camera. At the same time the CCTV recording will started. The subject was asked to stand and facing the camera starting from the distance of 3 meters from the camera. Then the subject will move to the 2nd position on 4 meters from the camera. Figure 4 shows on how the position of the subject increased by each meter.

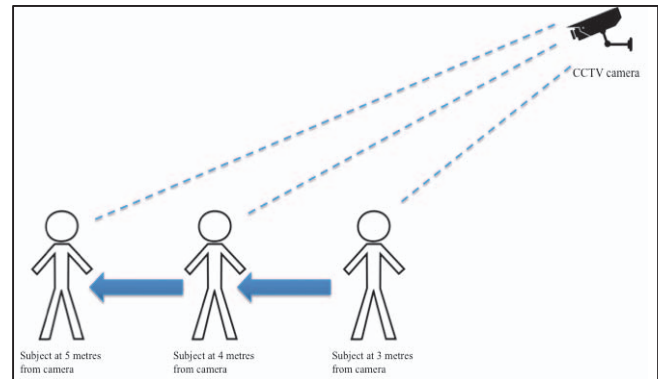


Fig. 4. Subject moves to different positions by increasing every meter

The subject repeated the process by increasing every 1 meter until 10 meters from the camera which was the last position. The process would be repeated again with lower resolution of camera. The same process applied to other subjects as well. Once all the resolution have been recorded we changed to the other camera for the same process. The frames would be extracted based on the position of subject

of each meter. Figure 5 is the one of examples of the extracted frames from each meter recorded by the CCTV.

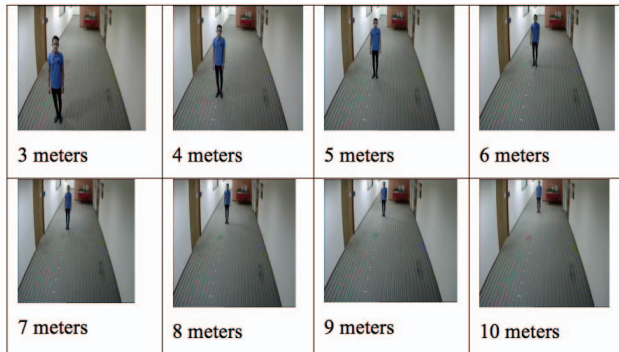


Fig. 5. Extracted frames by each meter recorded by CCTV

- *Extract face* – The faces were extracted manually from the frames. During the extraction we could see different size for the extracted face from each position and camera. Figure 6 is the example of the extracted face from the frames.



Fig. 6. Extracted face from the frame

- *Enhancement* – There were two methods for the enhancement part of our experiment: Enhancement using manual method and Enhancement module inside the Forensics Facial Identification System. The extracted faces were enhanced to improve the quality and results of the experiments.
- *Face recognition* – For this process, we used a system called ABIS® FaceExaminer from L – 1 Identity Solutions that is used by Cybersecurity Malaysia. This system can help security agencies and law enforcement to determine the identity through its face recognition feature. The comparison of the faces is done by searching from the database. The local gallery database is loaded during Enrolment phase. It also has

the face extraction modules and enhancement modules for the image. However in this research, we manually extracted faces from the CCTV recordings. Figure 7 is the main interface of the system.

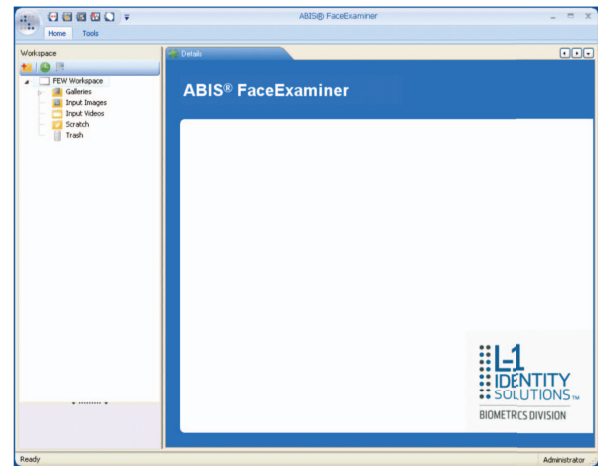


Fig. 7. The main interface for the ABIS® FaceExaminer

- *Result* – For this system the result of the face recognition is based on the scoring method. If the score of the result is 20 then the probability is $1 \times [10]^{-20}$ test which represent the probability of the tested face to false match. Confidence level is increasing if the matching score is higher. To understand the result of the system it is divided into 4:
 - i. The first rank – the probability of the first person in the list of matches is the same person. This precision and recall theory can be divided into two: true positives and false positives. For this research by using the dataset from the Enrolment phase we will find the best 2 from the 10 subject as the first rank.
 - ii. The lower rank - False negative. The result also can be assisted by the demographic of the subject such as gender, race, etc.
 - iii. No Match – The quality of the face is poor and the system are unable to provide the best feature extraction for the face detection

- iv. No fiducial features detected – The system are unable to run the auto face detection due to the poor quality of the face in the image.
- *Analysis* – We would look into the scoring of result obtained from the ABIS® FaceExaminer system. The comparison of scoring would be based on the types of camera, the resolution, and the distance of the subject from the camera. Based on the result, we could propose the suitable threshold of the camera in CCTV Forensics Analysis.

We have selected three resolutions for each camera in order to build up the data set. Table IV shows the list of selected resolution for the experiment.

TABLE IV. RESOLUTION SIZE OF EACH CAMERA

Types of Camera	Mobotix AllRound Dual M15 Day Night 90° Camera	ACTI E62	Raven RPS-AE1
Resolution			
Resolution 1	1024 x 768	800 x 600	704 x 299
Resolution 2	1280 x 960	1280 x 720	704 x 576
Resolution 3	1536 x 1024	1920 x 1080	960 x 576

IV. RESULTS AND DISCUSSION

Throughout the experiment, we found that different types of camera yielded different results. Each subject, camera, resolution, and distance may even produce negative results. We try to narrow down by focusing on specific subject, resolution, and distance in order to provide a comprehensive results.

We have selected one subject to show the relation of Distance versus Score.

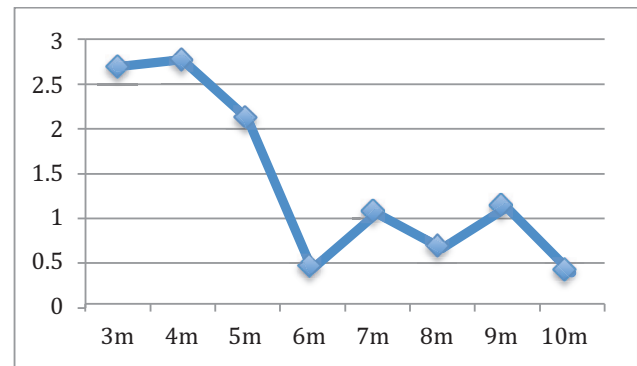


Fig. 8. Distance vs Score of Subject 1

Figure 8 shows the relation of score for one subject, using camera ACTI E62 with resolution 1920 x 1080. We can see the score is good at first three meters, but after distance of five meters from the camera, the scores become inconsistent. This is due to the increasing of distance of the subject from the camera.

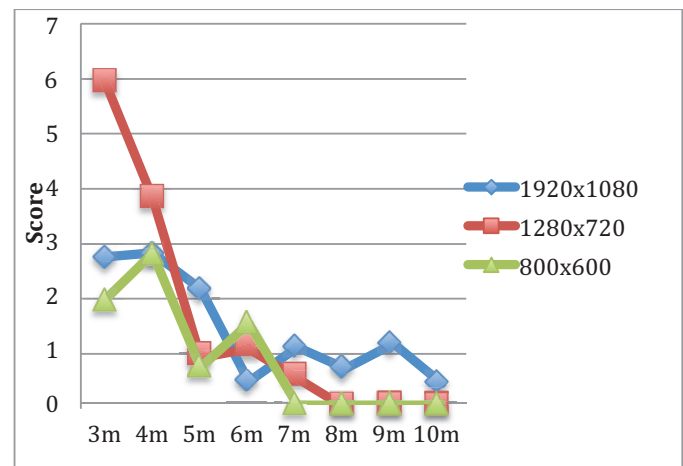


Fig. 9. Distance vs Score for 3 types of resolution using ACTI E62 camera

Figure 9 shows the score of one subject using camera ACTI E62 with three different resolutions. Although the resolution of 1280x720 is the highest score at 3 meters of distance, but after 7 meters, the facial tools were unable to detect the subject face for the resolutions of 1280x720 and 800x600. The resolution of 1920x1080 still produced acceptable score values for each distance. Table V shows the score results for each resolution.

TABLE V. SCORE RESULTS FOR DIFFERENT RESOLUTIONS USING CAMERA ACTI E62

	1920x1080	1280x720	800x600
3m	2.69	5.95	1.91
4m	2.77	3.82	2.76
5m	2.11	0.91	0.69
6m	0.43	1.07	1.48
7m	1.05	0.54	No detection
8m	0.66	No detection	No detection
9m	1.12	No detection	No detection
10m	0.39	No detection	No detection

V. CONCLUSION

We found out from the experiment that the resolution of camera and the distance between the subjects and camera will give effect to the results of forensics facial recognition analysis. Therefore, we proposed for the types of camera and resolution that have been tested, the acceptable distance between the subject and the camera should be less than 5 meters.

The limitation of this research is the fixed position of the camera. By having multiple positions, it may produce better quality of results. This can be included in future works, together with a framework that can assist the process of facial recognition for forensics analysis.

ACKNOWLEDGMENT

We would like to thank Ministry of Science, Technology, and Innovation (MOSTI) through project TechnoFund TF 0813 C268 (K2) entitled “GPU Enhanced Robust Multi-Dimensional Facial Identification System for CCTV Evidence in Video Forensics Analysis”. We also would like to thank CyberSecurity Malaysia for their support of the research.

REFERENCES

- [1] A. K. Jain, B. Klare, and U. Park. “Face recognition: Some challenges in forensics”. *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, pp.726-733. 2011.
- [2] S. Chen, E. Berglund, A. Bigdeli, C. Sanderson, and B. C. Lovell. “Experimental analysis of face recognition on still and CCTV images”. *IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance* 317-324. 2008.
- [3] S. Chen, and B. C. Lovell. “Illumination and expression in-variant face recognition with one sample image”. *Proceedings of the 17th International Conference on Pattern Recognition* 1, pp.300-303. 2004.
- [4] Z. Wang, A. C. Bovik, and L. Lu. “Why is image quality assessment so difficult?” *IEEE International Conference on Acoustics, Speech, and Signal Processing* 4 pp.3301-3313. 2002.
- [5] Z. Wang, E. P. Simoncelli, and A. C. Bovik. “Multiscale structural similarity for image quality assessment”. *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers* 2, pp.1398-1402. 2003.
- [6] B. Watson, J. Hu, and J. F. McGowan. “Digital video quality metric based on human vision”. *Journal of Electronic Imaging* 10(1) pp.20–29. 2001.
- [7] T. N. Pappas, and R. J. Safranek. “Perceptual criteria for image quality evaluation”. *Handbook of Image and Video Processing*. Academic Press. 2000.
- [8] L. Guo, and Y. Meng. “What is wrong and right with MSE?” *Eighth IASTED International Conference on Signal and Image Processing*, pp.212–215. 2006.