On using High-Definition Body Worn Cameras for Face Recognition from a Distance

Wasseem Al-Obaydy and Harin Sellahewa

Department of Applied Computing, University of Buckingham,
Buckingham, MK18 1EG, UK
{wasseem.alobaydy,harin.sellahewa}@buckingham.ac.uk
http://www.buckingham.ac.uk

Abstract. Recognition of human faces from a distance is highly desirable for law-enforcement. This paper evaluates the use of low-cost, high-definition (HD) body worn video cameras for face recognition from a distance. A comparison of HD vs. Standard-definition (SD) video for face recognition from a distance is presented. HD and SD videos of 20 subjects were acquired in different conditions and at varying distances. The evaluation uses three benchmark algorithms: Eigenfaces, Fisherfaces and Wavelet Transforms. The study indicates when gallery and probe images consist of faces captured from a distance, HD video result in better recognition accuracy, compared to SD video. This scenario resembles real-life conditions of video surveillance and law-enforcement activities. However, at a close range, face data obtained from SD video result in similar, if not better recognition accuracy than using HD face data of the same range.

Keywords: HD video, Face Recognition, Face Database, Surveillance, Eigenfaces, Fisherfaces, Wavelet Transforms

1 Introduction

Automatic recognition of human faces from video sequences has many applications. Most notable among them include law-enforcement, surveillance, forensics and content-based video retrieval. Much progress has been made in developing systems to recognise faces in controlled, indoor environments. However, accurate recognition of human faces in unrestricted environments still remains a challenge [10]. This is due to significant intra-class variations caused by changes in illumination, head pose and orientation, occlusion, sensor quality and video resolution [8, 9].

Normally, video signals captured by digital imaging devices are digitised at resolution levels lower than that of still images; hence the quality of a frame extracted from a video sequence is lower than that obtained from a still imaging device. Developing a robust video-based face recognition system that operates in unrestricted environments is a difficult task. This is due to the poor quality of face images in terms of image degradation, motion blur and low resolution. Therefore, the resolution of video frames could play a vital role in face

recognition from a distance. The understanding of the gains and losses of using high-resolution video in face recognition is an important factor when designing a biometric system to recognise faces in unrestricted conditions.

Recently, high definition (HD) video has been introduced as a new video standard that provides high quality video with high resolution as opposed to low-resolution standard definition video (SD). The availability of low-cost, miniature, high-definition video capture devices, combined with advanced wireless communication technologies, provide a platform on which real-time biometric systems that could recognise faces in unrestricted environments can be realised. The expectation is that, the recognition accuracy can be improved by increasing the video resolution. Recent studies have shown that using high quality/resolution video result in better face recognition accuracy [1, 14, 10]. Law-enforcement, forensics, video surveillance and counter-terrorism are areas that can benefit from such biometric systems. An example scenario is the real-time analysis a of video stream, captured by a camera worn on the uniform of a police officer to identify if a missing (or wanted) person is in the area that the police officer is patrolling.

This paper contributes to the current research in face recognition by investigating the use of HD body worn cameras to recognise faces from a distance and in outdoor conditions. The study looks at recognising faces captured at four different distance ranges in indoor and outdoor recording conditions. We evaluate the effects of using HD and SD video images for three benchmark face recognition algorithms: 1) Eigenfaces [15] 2) Fisherfaces [2] and 3) Wavelets [11]. A new face video database has been recorded at the University of Buckingham[1]. Videos of 20 subjects were acquired in HD and SD formats using a low-cost HD body worn digital video camera. An evaluation protocol is defined for the experiments conducted in this phase of the study.

The rest of the paper is organised as follows: Sec 2 introduces the features of the newly acquired HD/SD video database. Section 3 describes the three baseline face recognition algorithms used in this evaluation. Experiments and results are discussed in Sec. 4. Our concluding remarks and future works are presented in Sec 5.

2 High and Standard Definition (HSD) Video Database

2.1 High Definition vs. Standard Definition Video

The formats of NTSC, PAL and any video with vertical resolution less than 720 pixels are classified as standard definition (SD) video formats. Originally, NTSC and PAL are analogue standards, the digital representations of which can be obtained by digitising (sampling) the video frames. The NTSC video frame is digitised to 640×480 pixels, while a PAL video frame is sampled to 768×576 pixels [4]. Both NTSC and PAL systems have a 4:3 aspect ratio, and follow the

[1] The UBHSD database can be obtained for research purposes by contacting the second author of this paper

interlaced scanning system. The actual frame rate of NTSC video is 29.97 fps but it is often quoted as 30 fps, whereas the frame rate of PAL video is 25 fps [4].

In recent years, an increasing demand for high quality video has resulted in rapid adoption of HD digital video, particularly for home entertainment and digital TV broadcast. HD video is any video that contains 720 or more horizontal lines of the vertical resolution of the video frame. The Advanced Television System Committee (ATSC) states that the frame size of the HD video is either 1280×720 or 1920×1080 pixels [3]. All HD video formats support a widescreen aspect ratio of 16:9. Thus, HD video provides a high quality picture with high spatial resolution compared to the SD video. HD video with 720 pixels supports only progressive scanning, and is denoted by 720p, while HD video with 1080 pixels supports both interlaced and progressive scanning, and is denoted by 1080i and 1080p respectively [4]. Unlike SD video, HD video offers a variety of frame rates: 24, 30 and 60 fps.

2.2 HD body worn camera

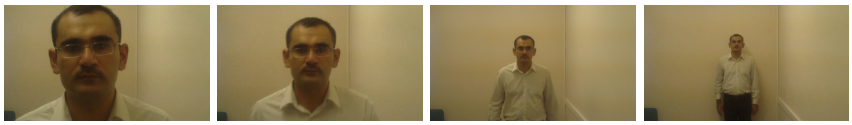
The videos in the database were acquired using a iOPTEC-P300; a HD body worn digital video camera, designed for police forces and security agencies for covert/overt surveillance and to collect real-time audio/video evidence. In order to maintain consistency of different physical properties of a camera (e.g. camera optics, lens) that affects its video quality, the same HD camera was used to capture both HD and SD videos. The SD video was recorded at a resolution 848×480 pixels, and at 25fps. The HD video was acquired at a resolution of 1920×1080p pixels, and at 30fps. Both the SD and HD videos were recorded in MOV file format.

2.3 UBHSD Database

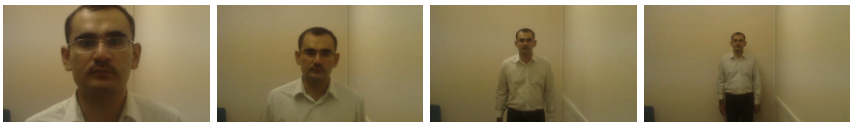
Data Collection The database contains a total of 160 videos of 20 distinct subjects. The videos of each subject were recorded in two sessions; each session includes two conditions: indoor and outdoor. The period between the two recording sessions was at least two days. In each condition, two video recordings (one HD and one SD) of the subject were captured sequentially by the same HD camera. All indoor recordings were captured in the same room under semi-controlled lighting with a uniform background. Outdoor videos were captured in an uncontrolled environment. These recording conditions represent realistic scenarios under which applications of face recognition at a distance can be applied.

During a recording, a subject walk a distance of 4 meters (indoor) and 5 meters (outdoor) toward the camera, from a start-point to a stop-point, providing face data at different distances. The minimum distance betweeen the camera and the subject (stop-point) is a meter. The subjects face the camera while they walk toward it. However, they were free to walk in a natural way which included head movements and facial expressions. A video recording lasted about 5 to 10 seconds depending on the speed at which the subject walks. Figure 1 shows video

frames extracted from a typical indoor and outdoor recording conditions for a subject. The frames of HD and SD videos are scaled down at different levels for display purposes.



(a) HD video, Indoor (distance range from left to right: R_1 - R_4)



(b) SD video, Indoor (distance range from left to right: R_1 - R_4)

(c) HD video, Outdoor (distance range from left to right: R_1 - R_4)

(d) SD video, Outdoor (distance range from left to right: R_1 - R_4)

Fig. 1: A sample of indoor and outdoor video frames of the High/Standard Definition Video Database

Data preparation Twelve frames of from each video are selected in a systematic way to capture the subject at four distance ranges from the camera position. Each distance range is represented by 3 frames. The frames in the first range, *Range 1* (R_1), are nearest to the camera, while the frames in the fourth range, *Range 4* (R_4), are the farthest away from the camera. Each row in fig. 1 consists of four frames, each representing a distance range.

The total walking distance is sectioned into 4 ranges is by dividing the total number of video frames by 4. Then, the mid, $mid + 5$, and $mid + 10$ frames in each range are selected and extracted from the video. This ensures that the subject, who appears in HD frames at certain distance range, appears also in their corresponding SD frames at the same distance range from the camera. In some cases, the mid+15 frame was chosen instead of one of the three frames when

(a) HD Indoor, Range 1 - 4 (b) HD Outdoor, Range 1 - 4

(c) SD Indoor, Range 1 - 4 (d) SD Outdoor, Range 1 - 4

Fig. 2: Examples of cropped and rescaled face images from HD and SD videos captured in indoor and outdoor conditions

the latter suffers from severe motion blur. Yet, the database consists of blurred face images, faces with eyes closed and slightly varying poses. Each subject has 96 face images, thus the total number of face images in the database is 1920.

The face region in each frame was manually cropped at the top or middle of the forehead, bottom of the chin, and at the base of the ears. Then, all face images were converted to grayscale and rescaled to size of 128×128 pixels. The experiments reported here are based on these images. Figure 2 shows the cropped and rescaled face images extracted from the respective HD and SD videos frames in Fig. 1.

3 Baseline algorithms

A brief description of each of the three benchmark face recognition algorithms, namely Eigenfaces, Fisherfaces and wavelet-based face recognition is given in this section. As shown in Fig. 1, videos of subjects in UBHSD database are captured under varying lighting conditions. There are many normalisation techniques that can be used to deal with the problem of varying illumination conditions [13, 7]. We tested the effect of the commonly used histogram equalisation (HE) and z-score normalisation (ZN) on the recognition rates of the three algorithms. For Eigenfaces and Fisherfaces, ZN was applied on the cropped and rescaled face images while for the wavelet-based scheme, the selected wavelet subband was normalised by ZN.

3.1 Eigenfaces

Turk and Pentland [15] presented the Eigenfaces approach using Principle Component Analysis (PCA) to efficiently represent face images. PCA is a statistical analysis tool used to reduce the large dimensionality of data by exploiting the redundancy in multidimensional data. In this approach, each face image in the

high dimensional *image space* can be represented as a linear combination of a set of vectors in the new low dimensional *face space*. These vectors, calculated by PCA, are the eigenvectors of the covariance matrix of the face images in the training set. Each eigenvector can be displayed as a "ghostly" face image, hence eigenvectors are commonly referred to as eigenfaces. When a probe face image is presented for recognition, it is projected into the *face space* and a nearest neighbour classification method is used to assign an identity to the probe image.

3.2 Fisherfaces

Belhumeur et al. [2] presented Fisherfaces, a face recognition scheme claimed to be insensitive to illumination variations and facial expressions. The authors state that since the training images are labeled with classes (i.e. individual identities), it makes sense to exploit class information to build a reliable method to reduce the dimensionality of the feature space. This approach is based on using class specific linear methods for dimensionality reduction and simple classifiers to produce better recognition rates than Eigenfaces method which does not use the class information for dimensionality reduction. Fisher's Linear Discriminant Analysis (FLD or LDA) is used to find a set of projecting vectors (i.e. weights) that best discriminate different classes. FLD achieves that objective by maximising the ratio of the between-class scatter to that of the within-class scatter.

3.3 Wavelet-based face recognition

Discrete wavelet transforms (DWT) can be used as a dimension reduction technique and/or as a tool to extract a multiresolution feature representation of a given face image [5, 11, 6]. In the enrolment stage, each face image in the gallery set is transformed to the wavelet domain to extract its facial feature vector (i.e. a subband). The choice of an appropriate subband could vary according to the operational circumstances of the recognition application. The decomposition level is predetermined based on the efficiency and accuracy requirements and the size of the face image. In the recognition stage, a nearest neighbour classification method is used to classify the unknown face images.

4 Experiments and Results

In this paper, we report the results of the first phase of our evaluation of HD and SD video in face recognition from a distance. Firstly, we define an evaluation protocol for the HSD video database to ensure repeatability and comparability of the work reported here and for future works using this database. The evaluation protocol is introduced in Sec. 4.1 followed by experimental results in Sec. 4.2.

4.1 Evaluation protocol

The evaluation protocol involves four configurations for each video resolution: 1) Matched Indoor (MI), 2) Matched Outdoor (MO), 3) Unmatched Indoor (UI)

and 4) Unmatched Outdoor (UO). Each configuration has four test cases (e.g. $MI_1, \ldots, MI_4$,). The gallery set G of test-case i ($i = 1, \ldots, 4$) consists only range R_i face images in Session 1. For each test case, images from all four ranges, in both indoor and outdoor videos in Session 2 are used as probe images (P). There is no overlap between the gallery and probe sets. In Matched configurations, both the gallery and probe images come from the same video resolution. In Unmatched configurations, gallery and probe images are from different video resolutions. For each test-case, the gallery set consists of 60 images (3 images per subject) and the probe set consists of 480 images (24 images per subject). Table 1 describes the gallery and probe sets for different configurations.

Table 1: The test configurations for the HSD video database

Configuration		Session 1				Session 2			
		HD video		SD video		HD video		SD video	
		Indoor	outdoor	Indoor	outdoor	Indoor	outdoor	Indoor	outdoor
HD	MI_i	G,R_i				P,R_{1-4}	P,R_{1-4}		
	MO_i		G,R_i			P,R_{1-4}	P,R_{1-4}		
	UI_i	G,R_i						P,R_{1-4}	P,R_{1-4}
	UO_i		G,R_i					P,R_{1-4}	P,R_{1-4}
SD	MI_i			G,R_i				P,R_{1-4}	P,R_{1-4}
	MO_i				G,R_i			P,R_{1-4}	P,R_{1-4}
	UI_i			G,R_i		P,R_{1-4}	P,R_{1-4}		
	UO_i				G,R_i	P,R_{1-4}	P,R_{1-4}		

4.2 Recognition results

A number of experiments have been conducted using the newly created HSD video database to evaluate the use of HD and SD video in face recognition from a distance. All three face recognition algorithms use L1 (CityBlock) distance to calculate a match score between two feature vectors. The Haar wavelet transform is used for the wavelet-based recognition and we report results for LL_3 and LH_3 subbands based on recent work in [12]. Rank one recognition accuracy for MI and MO configurations based on Eigenfaces (PCA), Fisherfaces (LDA) and DWT (LL-subband and LH-subband) are presented in Fig. 3 through Fig. 6. We also report results for UI configuration, based on LH_3 subband features (with z-score normalisation), in Tab. 2.

The overall recognition rates of all test cases indicate that the use of HD video data for face recognition at a distance has a significant advantage over that of SD video data. This observation is in agreement with our expectation that using high-resolution video data would lead to better recognition rates for face recognition at a distance. However, a closer examination of individual tests

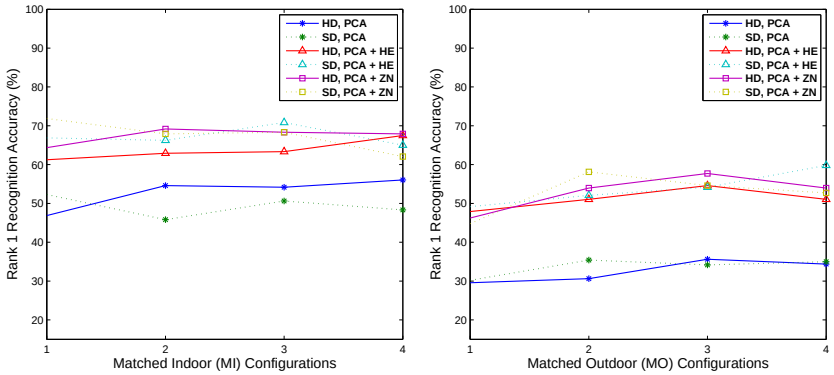


Fig. 3: Rank 1 recognition accuracy of HD & SD video using PCA

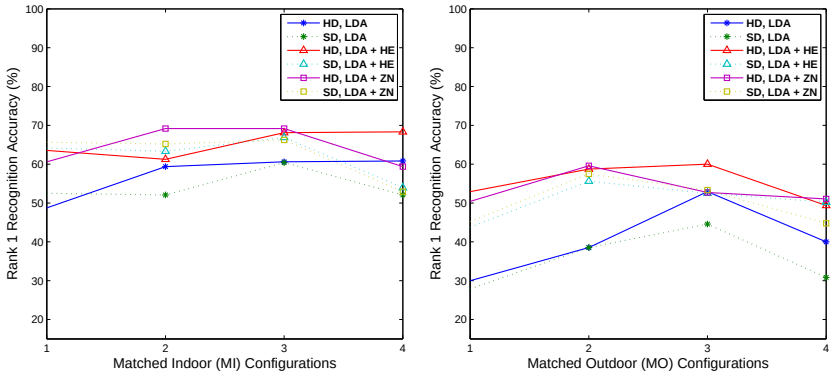


Fig. 4: Rank 1 recognition accuracy of HD & SD video using LDA

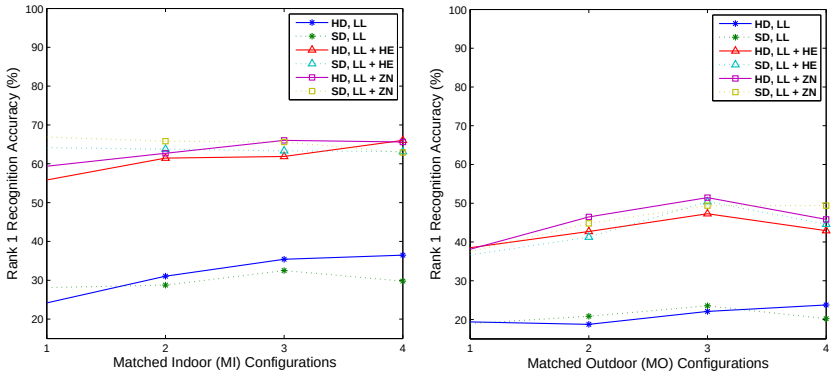


Fig. 5: Rank 1 recognition accuracy of HD & SD video using LL_3

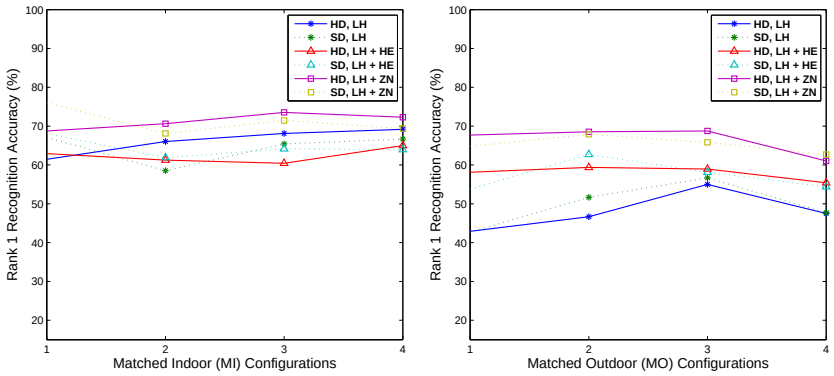


Fig. 6: Rank 1 recognition accuracy of HD & SD video using LH_3

reveals an interesting pattern. When the gallery set is the collection of face images nearest to the camera (i.e. Test $Case_1$), SD video data result in similar if not significantly higher recognition accuracy compared to that of using HD video data, irrespective of the distance range of the probe images. There could be a number of reasons for this behaviour.

Firstly, a Range 1 face image taken from HD video has to be down sampled by a much larger factor than the one used for a face image taken from SD video (to produce a 128×128 pixel face image). The resulting degradation of quality depends on the down sampling technique (in our case, we used MATLAB 'imresize' with the default bicubic interpolation) and it is higher on face images taken from HD videos than it would on face images taken from SD videos. Aliasing, caused by down sampling, could also be a factor. To establish if there downsampling has an adverse effect on HD video images at Range−1, we repeated the Matched Indoor tests using Gallery images from $Range_1$ for different face sizes: 1) 64×64, 2) 96×96, 3)160×160 and 4) 200×200. The rank 1 recognition accuracies for HD and SD data are given in Tab. 3. The results give some indication that less downsampling is better for HD (the size of face images captured in HD at close range are much larger than those captured from SD). This requires further investigation to identify why at Range1 SD-SD outperforms HD-HD.

On the other hand, it could be that "more is less", meaning having too much information (e.g. high image resolution) is not necessarily a good thing in face recognition. This could be the reason for lower accuracy of HD video image using Eigenfaces approaches. It is also possible that 60 high resolution training samples (3 per subject) are insufficient to obtain a good discriminative *face space* for recognition because of data redundancy. We noticed a significant increase in recognition accuracy when the number of training samples was increased from 1 to 3. Note that the training images used for each subject are obtained from video frames that are nearer to each other. Hence, there is little variation among them. This is in contrast to the gallery data selection techniques proposed in [14], which

aim to use training samples that capture variations. In our test configurations, we try simulate conditions that may have a limited choice of gallery images for each subject.

We have reproduced in Tab. 4, a selection of experimental results by Thomas et al. in [14] that shows the recognition accuracy of three different cameras. The JVC is a high-definition camera and the Canon is standard-definition camera. Note that in [14], the number of samples used in the gallery set for the selected results is 12 or 15 as opposed to 3 samples we have used in our evaluation.

Figure 3 through Fig. 6 also present rank one recognition rates for two illumination normalisation techniques. Normalisation has significantly improved the recognition rates of all algorithms. Its effect is prominent in Eigenfaces and LL-subband based recognition; two feature representations that are known to be severely affected by varying lighting conditions. In terms of HD video vs. SD video in face recognition, the HD video is still the better of the two standards, except when gallery images are from Range 1, SD video is the better option. Surprisingly, z-score normalisation resulted in much higher recognition accuracy than the commonly used histogram equalisation for illumination normalisation.

A comparison of the three face recognition algorithms shows that the recognition rates of Fisherfaces approach is similar, if not better than the recognition rates of Eigenfaces approach. However, simply using the LH-subband of wavelet transformed images as face features significantly outperforms both Fisherfaces and Eigenfaces schemes.

It is worth noting the significant decrease in recognition accuracy when outdoor video images are used as a gallery set. These results highlight the challenges of recognising faces from a distance and in unrestricted environments.

Table 2: Rank 1 recognition accuracy of Matched and Unmatched configurations

Gallery Set	Probe Set	Gallery Image Range			
		Range_1	Range_2	Range_3	Range_4
HD	HD	68.75	70.62	73.54	72.29
HD	SD	68.75	65.83	72.08	75.00
SD	SD	76.04	68.12	71.46	69.38
SD	HD	75.83	71.04	71.25	68.33

5 Conclusions and Future Work

In this paper, we presented a performance evaluation of HD and SD video in face recognition from a distance. We created a new face biometric database consisting HD and SD videos of 20 different subjects, captured at different distances using a low-cost HD body worn camera. We used three benchmark algorithms, namely

Table 3: Recognition accuracy vs. face size. Gallery Images from Range_1

	Face image size (pixels)				
Gallery/Probe	64×64	96×96	128×128	160×160	200×200
HD/HD	68.12	72.08	68.75	72.92	69.58
SD/SD	75.42	76.46	76.04	74.79	75.00

Table 4: Rank 1 recognition rates by Thomas et al. in [14], Tab. 18.1

Gallery (NEHF)	Probe	Accuracy Rate	Number of Images
JVC	JVC	82.9	12
JVC	Canon	78.1	15
Canon	Canon	79.0	12
Canon	JVC	76.2	12

Eigenfaces, Fisherfaces and Wavelets-based approaches for the evaluation of HD and SD video in face recognition from a distance.

The overall recognition rates of all test configurations favour the use of HD video data for face recognition from a distance as opposed to using SD video data. This is in line with the expectation that high-resolution video data would lead to better recognition rates for face recognition from a distance. Previous work also suggests the same. However, for recognition at a close range, HD video might not provide an added benefit in terms of recognition accuracy, when compared with SD video.

This brings us to the important question; should we use HD video or SD video for face recognition from a distance? Based on the evaluation presented here, the choice of HD or SD depends on the quality of the gallery set and the probe images presented for identification. For applications where person identification from a distance is a requirement, HD video offers a clear advantage over SD video. However, SD video has shown to produce higher recognition rates for face recognition at a close range. Therefore, a face recognition system in unrestricted environments (e.g. CCTV with automatic face recognition) should be able to select the appropriate resolution (or zoom in and out) when attempting identify a person.

It must be emphasised that the benefits of HD video comes at the cost of high bandwidth, storage and processing requirements. In situations where the use of HD video is unaffordable, super resolution techniques could be used to improve the accuracy of low-resolution, SD video data. It is also important to understand the effects of various pre-processing techniques (e.g. resizing, illumination normalisation) that are commonly applied on face images prior to using them as gallery or probe images. These are important questions that require

further investigation. This brings us to the next phase of the evaluation. Our future works include the use and evaluation of super resolution techniques in face recognition at a distance. We will also evaluate the performance of state-of-the-art face recognition algorithms on the newly acquired HD and SD video database and investigate the performance of HD video data with varying sample sizes in the gallery set.

References

1. Bailly-Bailliére, E., Bagnio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariéthoz, J., Matas, J., Messer, K., Popovici, V., Porée, F., Ruiz, B., Thiran, J.: The BANCA Database Evaluation Protocol. In: Audio-and Video-Based Biometric Person Authentication. pp. 625–638. Proc. 4th Int'l Conf. AVBPA (June 2003)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (July 1997)
3. Browne, S.E.: High Definition Postproduction: Editing and Delivering HD Video. Focal Press (December 2006)
4. Chapman, N., Chapman, J.: Digital Multimedia, 3rd ed. John Wiley & Sons, Ltd. (February 2009)
5. Chien, J.T., Wu, C.C.: Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(12), 1644–1649 (December 2002)
6. Ekenel, H.K., Sankur, B.: Multiresolution face recognition. Image and Vision Computing 23(5), 469–477 (May 2005)
7. Gross, R., Brajovi, V.: An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. In: Audio-and Video-Based Biometric Person Authentication. pp. 11–18. Proc. 4th Int'l Conf. AVBPA (June 2003)
8. Kung, S.Y., Mak, M.W., Lin, S.H.: Biometric Authentication:A Machine Learning Approach. Prentice Hall, New Jersey (2005)
9. Park, U.: Face Recognition: face in video, age invariance, and facial marks. Ph.D. thesis, Michigan State University, USA (2009)
10. Phillips, P.J., Flynn, P.J., Beveridge, J.R., Scruggs, W.T., O'toole, A.J., Bolme, D.S., Bowyer, K.W., Draper, B.A., Givens, G.H., Lui, Y.M., Sahibzada, H., Scallan, J.A., Weimer, S.: Overview of the multiple biometrics grand challenge. In: International Conference on Biometrics. pp. 705–714. Proc. (June 2009)
11. Sellahewa, H., Jassim, S.: Wavelet-based face verification for constrained platforms. In: Biometric Technology for Human Identification II. Proc. SPIE, vol. 5779, pp. 173–183 (March 2005)
12. Sellahewa, H., Jassim, S.: Image quality-based adaptive face recognition. IEEE Transactions on Intrumentation & Measurements 59, 805–813 (April 2010)
13. Shan, S., Gao, W., Cao, B., Zhao, D.: Illumination Normalization for Robust Face Recognition Against Varying Lighting Conditions. IEEE International Workshop on Analysis and Modeling of Faces and Gestures pp. 157–164 (2003)
14. Thomas, D., Boyer, K.W., Flynn, P.J.: Strategies for improving face recognition from video. In: Ratha, N.K., Govindaraju, V. (eds.) Advances in Biometrics - Sensors, Algorithms and Systems, chap. 18, pp. 339–361. Springer (2008)
15. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)