Automated Selection of Optimal Frames in NIR Iris Videos

Nitin K. Mahadeo, Andrew P. Papliński, Sid Ray
Clayton School of Information Technology
Monash University
Email: {Nitin.Mahadeo, Andrew.Paplinski, Sid.Ray}@monash.edu

Abstract—A relatively new trend in the iris biometric area is the use of videos as a capturing device. Frame by frame approach is richer in information and gives more flexibility as opposed to the use of traditional still images. However, the quality, shape and size of the iris may vary from one frame to another. In this paper, we propose a new technique for selecting the best frames in an iris video. Taking advantage of the temporal correspondence in iris frames, we classify iris videos into 3 categories, namely Adequate, Motion Constrained and Time Constrained. Frames with blinks and off-angle gaze are eliminated using frame averaging and correlation. Quality factors, namely motion blur, out of focus, translational motion and lighting present in iris videos are detected and their effect on recognition performance is investigated. Experimental results are carried out on both the MBGC NIR Iris Video and the MBGC NIR Iris Still datasets from the National Institute for Standards and Technology (NIST). Firstly, this work demonstrates that the proposed optimal frame selection technique in NIR Iris Videos leads to significant improvement in recognition performance. Secondly, the performance of NIR Iris Still images vs. NIR Iris Videos is compared. Thirdly, we show that interoperability between iris frames and iris images in an iris recognition system affects performance. Finally, the computational time and the elimination of noisy frames at each stage using the proposed method are examined.

I. INTRODUCTION

As iris biometric matures, there has been a recent shift from still-images to video-images. While this approach in iris biometric relaxes some of the constraints encountered in still-images it also brings along new challenges which need to be addressed. A still image is not always optimal in terms of quality and often contain noise such as reflections, specular highlights and occlusions [1]–[3]. Several image enhancement techniques have been proposed over the years to improve the quality of the iris image. On the other hand, iris videos offer more flexibility i.e., there are more frames to choose from instead of a few ones. However, in the case of iris videos, some frames might be out of focus and severely occluded while others might be sharp and of high quality. Moreover, frame processing also calls for a more efficient approach as the number of frames is significantly higher. This paper presents a technique aimed at reliably processing iris videos such that the best quality frames are selected. This has several advantages. Firstly, by choosing the best iris frames, the chances of a successful segmentation are increased. Secondly, this also decreases chances of a subject having to resubmit his or her eye for image capture. Thirdly, segmentation and encoding of high quality frames also translates into better performance as shown in our experiments.

A. Background and Motivation

The vast majority of work carried out in the iris biometric field has been conducted on iris still images [3]. In [4], Kalka *et al.* investigate the effect of a wide range of factors on recognition performance using still images. In [5], Zuo and Schmid extend the work carried out in [4] to both global and local image quality assessment. These methods require the iris region to be segmented prior to image quality assessment evaluation making them computationally intensive and unsuitable for real time applications, especially in the case of Near-infrared (NIR) iris videos where hundreds of frames are being used. On the other hand, there has been relatively less work in the iris video area. In [6], [7], images containing blink and insufficient iris data are manually replaced. In [8], the best quality images are selected using "Automatic Image Quality Measurement Algorithm" (AIQM), an image quality assessment previously used in face recognition systems and "Human Image Quality Measurement"(HIQM) where the best frames are selected by visual inspection. In [9], Zhou *et al.* propose the use of interpolation and detection of the presence of specular reflections and glare to discard frames containing interlacing, blink and blur.

In the literature, image quality evaluation is typically carried out based on a single or pair of criteria [6], [10], [11]. In this paper, a different approach is presented. Users are first classified into a category depending on the quality of the iris videos. Noisy frames are then eliminated in an efficient manner. Blink and off-angle frames are first discarded. In the next step, frames which are out of focus are removed using a Laplacian based operator and frames with motion blur are eliminated. Translational error and localized motion blur present in iris images are also investigated. Finally, frames with poor lighting and illumination are dismissed.

The remainder of this paper is organized as follows. Section II is a brief description of the datasets used and the different types of videos present in the database. Section III describes the procedure implemented to evaluate the quality of iris frames in the category of Adequate videos. Section IV introduces the reader to challenging video sequences in the dataset and our approach to dealing with them. Finally, we present and discuss our results in Section V and Section VI.
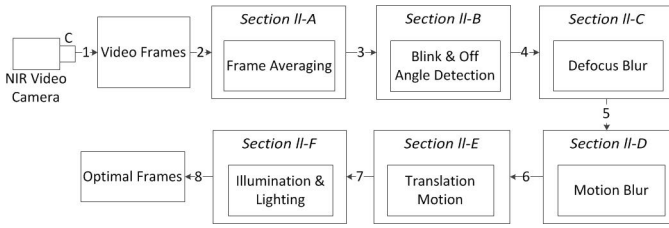
Fig. 1: A flowchart depicting the different stages of the frame selection procedure.

Fig. 2: Averaged frames (a) and (b) fall into the cooperative category as a clear global minimum can be detected in the central region.

II. MBGC Data & Video Sorting

Two of the recordings of the Multiple Biometric Grand Challenge (MBGC) made available by the National Institute for Standards and Technology (NIST) are used in this paper. Still iris images and iris video sequences were collected using an LG 2200 sensor. There are 290 videos in all in the NIR Video dataset with the number of frames ranging from 5 to 1018 frames per video. The clips are in MPEG-4 format and each video contains either the right or left eye only. The NIR Still dataset consists of 1770 images captured with an LG 2200 camera. It contains images with significant challenges in terms of image quality [12].

NIR Iris Videos are classified into three classes namely, Adequate, Motion Constrained and Time Constrained. Videos with a large number of frames in which the majority of iris regions are centrally positioned with limited user movement are considered to be Adequate. Videos containing lots of head movement are referred to as Motion Constrained and clips which contain a limited number of frames fall into the Time Constrained category. Over 95% of clips in the NIR Iris Video were found to belong to the Adequate category. Section IV deals with Time and Motion Constrained videos and the next Section discusses our approach to processing videos in the Adequate division.

III. Adequate Videos

In this section, frame averaging is performed on each video clip followed by blink and occlusion detection using correlation. Out of focus and images containing motion blur are detected. Translational motion and localized motion blur are also investigated. Finally, frames with illumination and lighting are eliminated. This is illustrated in the flowchart in Figure 1.

A. Frame Averaging

In order to distinguish between the three types of videos mentioned in Section II, the following approach is used. By computing the average of all frames in the iris video, one would expect that the averaged frame in an adequate video to contain a global minimum caused by the mean of the low intensity of the pupil. Two examples of such averaged frames are shown in Figure 2. Videos containing lots of head and eye movements have no such clear minima as the low pupil intensity is diluted by other higher intensity regions in the iris
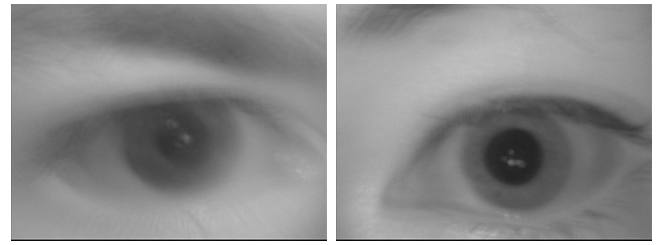
frames. As for clips with limited number of frames, they would be unreliable. The convolution between the averaged frame f_{avg} and that of a low pass filter, H, is computed as shown in expression (1) where $\otimes$ denotes the convolution operator.

$$J = f_{avg} \otimes H \tag{1}$$

The minimum of the convoluted output image, J in equation (1) is computed. This information gives us a rough approximation of the strength and location of the iris region in the averaged frame. Frames with a low minimum or whose iris are not centrally located are classified as Time Constrained or Motion Constrained depending on the number of iris frames. Hence a decision can be made into which category each iris video falls into.

B. Blink & Off-Angle Detection

The next step which follows deals with the elimination of frames containing blinks, off-angle and low iris visibility. Only frames in which the iris is centrally located i.e., those frames with minimal occlusion and adequate amount of iris information are retained. This approach also facilitates the segmentation process which is arguably the most important stage in an iris recognition system. The majority of frames have their pupil centrally located in an adequate video. Therefore, by computing the cross-correlation coefficient between the averaged frame and its individual frames, those frames which contain severe occlusion, off-angle and blinks can be detected as they have low correlation compared to the averaged frame. Thus, these frames can be detected and discarded. The correlation coefficient, r, is given by equation (2)

$$r = \frac{\sum_i \sum_j (f_{ij} - \bar{f})(A_{ij} - \bar{A})}{\sqrt{\sum_i \sum_j (f_{ij} - \bar{f})^2 \sum_i \sum_j (A_{ij} - \bar{A})^2}} \tag{2}$$

where $\bar{f}$ is the mean of a frame f_{ij} in a given video and $\bar{A}$ is the mean of the averaged frame A_{ij}. This process ensures that only the frames whose iris are centrally located with sufficient valid information are retained. Figure 3 shows some of the results of the proposed technique. The fist row contains images whose correlation coefficient, r, was found to be smaller than a given threshold and were therefore discarded. Frames in the second row have a high correlation and are retained for further
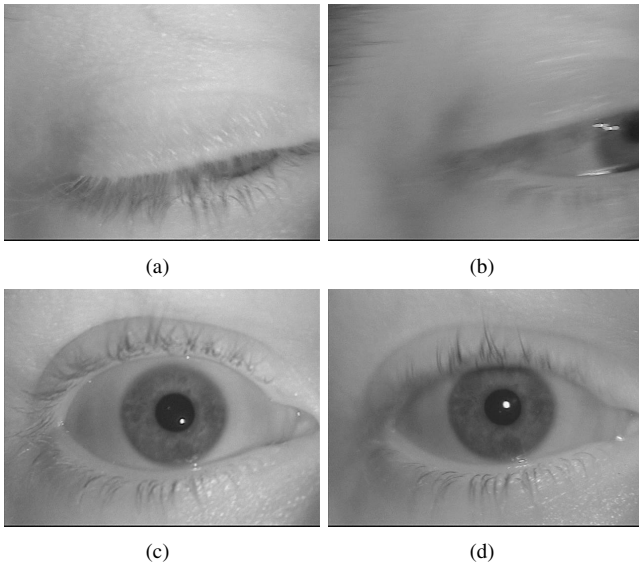
(a) (b)

(c) (d)

Fig. 3: The 4 Frames above are from the same video. Frame (a) and (b) were discarded as a result of low correlation coefficient while frames (c) and (d) are retained.

processing. In our experiments, the threshold value of r is set to 0.8.

C. Defocus Blur

Blur is typically represented as a shift-invariant model, where an image is convoluted with a point spread function (PSF), usually a Gaussian; the wider the PSF, the blurrier the image. One of the effects of blur is that high spatial frequencies in the image are abated. By measuring the power in an image obtained by filtering with a proposed convolution kernel, Daugman showed that a 2-D focus assessment can be made [10]. In the recent literature, several variants thereof have been proposed [4], [5]. Out-of-focus video frames usually affect the segmentation stage by causing the edges at the pupil and iris boundaries to become smoother. In this paper, a gradient based method similar to that proposed in [13] is implemented to detect blurred images. The Laplacian, being a second order isotropic operator, is suitable for the detection of high frequencies, analogous to strong edges in the images. The focus score, S, of a given frame is calculated by computing the variance of image Laplacian as shown in expression (3)

$$S = \sum_i \sum_j (\Delta f_{i,j} - \overline{\Delta f})^2 \qquad (3)$$

where $\overline{\Delta f}$ is the mean of the Laplacian of a given frame f and S is the focus score. For each video, the focus score is normalized between 1-100, only frames in the upper quartile are retained. The advantage of this approach compared to the one proposed in [10] is that it does not depend on a convolution kernel to determine the focus of an image but instead uses the image information itself. It should also be mentioned that frames often contain strong specular reflections but are still in

focus. A high pass filter is used to detect reflections present in the images. These strong gradients caused by reflections are ignored in our computations as they can bias the overall focus score.

D. Motion Blur

Proposed techniques in the literature focus on detecting motion blur in still images by comparing different regions of the iris. This can be computationally expensive as this approach requires iris segmentation. In this section we propose a computationally simple and more efficient way of detecting motion blur in iris videos. Following the removal of frames detected in Section III-B and Section III-C, the remaining consecutive frames are each considered as being part of a subgroup. For instance, frames before and after a blink in a clip are considered as two separate subgroups. Careful observation indicates that there is a change in iris position or amount of visible iris data following blinks, blur and occlusion. For example, in the case of blinks, the upper eyelid does not go back to its exact previous position. It may cover the iris region more or less after the blink occurs. This can also be the case in out of focus frames as a result of head movement. User inclination and localized motion blur are explored further in Section III-E. For now, we focus on a computationally fast and inexpensive motion blur detection technique. As we explore each subgroup, considering that the frame rate is 30 fps, this suggests that ideally there should not be much change from one frame to another within a given subgroup. However, contrary to out of focus images, motion blur affects the image locally, most frequently due to movement of the eyelid or the eyeball. The first row in Figure 4 shows two consecutive frames within a given subgroup containing motion blur and the second row shows another pair of adjacent frames with no motion blur.
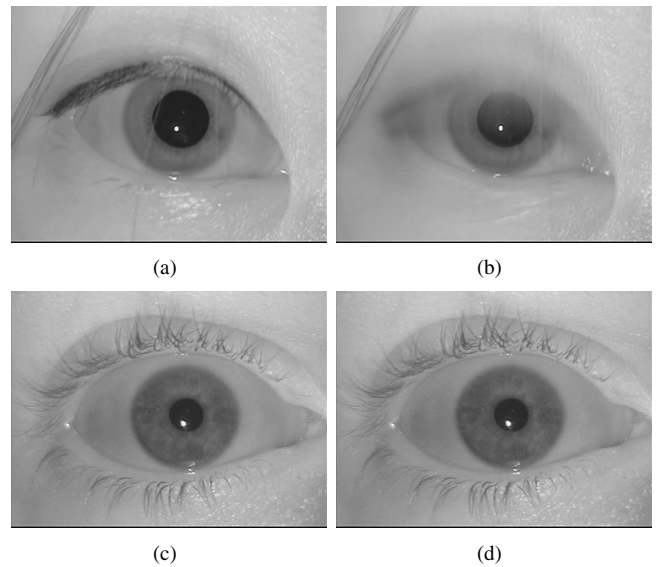

(a) (b)

(c) (d)

Fig. 4: (a) and (b) are adjacent frames containing motion blur. Consecutive frames (c) and (d) have no motion blur.
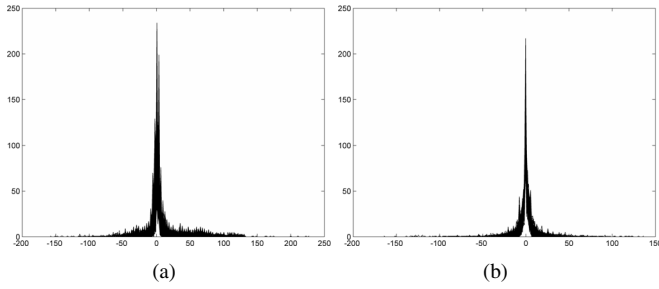
Fig. 5: (a) is the histogram of the differential frame of Figure 4(a) and 4(b) with a standard deviation of 18.9 (b) is the histogram of the differential frame of Figure 4(c) and 4(d) whose standard deviation is 7.0.

The histograms of their respective differential frames of these two pairs of consecutive frames are shown in Figure 5. The standard deviation of the histogram of the differential frame containing motion blur in Figure 5(a) is much higher than the one shown in Figure 5(b) without motion blur. This also shows that this error between two consecutive frames, as a consequence of motion blur in this case can be measured. A fast and computationally inexpensive way of detecting motion blur is achieved by computing the MSE error between adjacent frames within each group. The MSE is given by expression (4)

$$MSE = \frac{\|f_t - f_{t+1}\|^2}{N} \qquad (4)$$

where the squared norm of the difference between adjacent frames f_t and f_{t+1} is divided by the number of elements N. Following this rationale, it can be shown that adjacent frames containing motion blur have a higher MSE as opposed to stationary frames. The MSE of adjacent frames containing motion blur in Figure 4(a) and Figure 4(b) is found to be 388.1 while the MSE of frames in Figure 4(c) and Figure 4(d) is only 49.6. This appreciable difference between these two values is used to determine if motion blur is present. A threshold of 100.0 is used in our experiments.

E. Translational Motion

A high MSE was also recorded between adjacent frames in a given subgroup although no motion blur was present. Upon investigation, it was determined that this was due to object motion. This type of motion usually occurs due to eyeball and head movement, however there is no blur present in such images. In this section, we describe how we are able to differentiate between frames containing motion blur in the iris region and those with object motion only. The ability to evaluate how much the object has moved also gives rise to the possibility of aligning the current frames with the previous one. Moreover, this information facilitates the segmentation process as we can track the location of the iris region from frame to frame if required. As a result of the difference between the two objects of interest, the iris in this case, the error should ideally be small. Several motion estimation techniques such as Phase Correlation, Optical Flow and Block
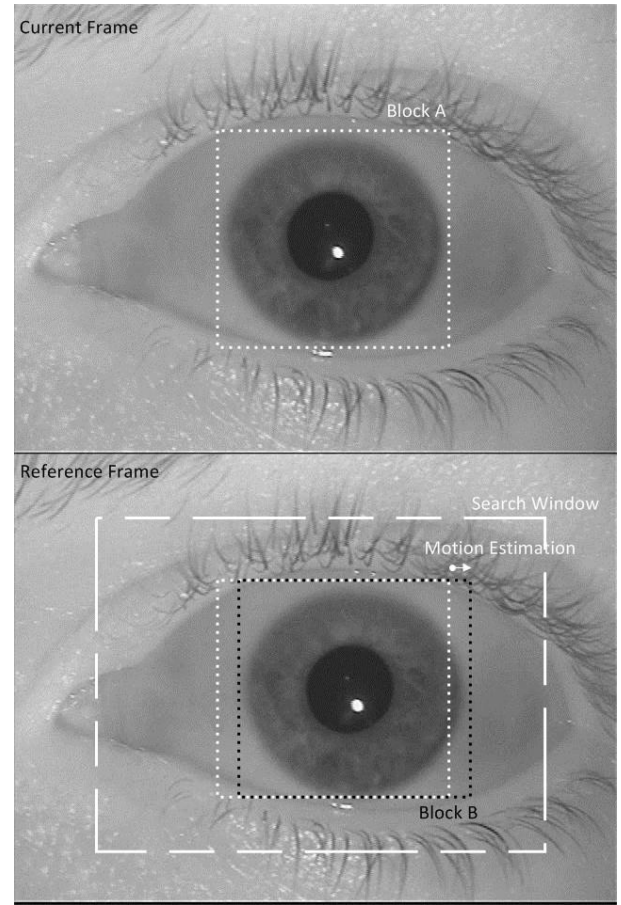

Fig. 6: Block A is the segment to be matched with other blocks of similar dimensions in the reference frame within the search window. Here, Block B is the matching block found with the smallest MSE and the motion vector is determined.

Matching techniques have been proposed in the literature. In this work, a block matching method is implemented using a minimum MSE criterion for consistency and comparison with the work in the previous section [14]. The assumption made here is that the object motion is only translational. From Section III-B, we are also aware that the remaining frames have their iris centrally positioned. Therefore, the translational motion can be estimated by setting the boundaries of the search window in a central position and blocks in the current and reference frames can be compared to estimate the motion within the region enclosed by the search window. This is illustrated in Figure 6 where Block A in the current frame is compared to other blocks of the same dimensions in the reference frame within the search window region. Here, Block B in the reference frame corresponds to the correct alignment and hence the motion vector is estimated.

The MSE between the current and reference frames in Figure 6 was found to be 242.0. The block matching technique reveals that the best matching block occurs in the spatial domain at the position of Block B where the MSE is equal to 22.8. Due to noise and lighting conditions, the MSE is not
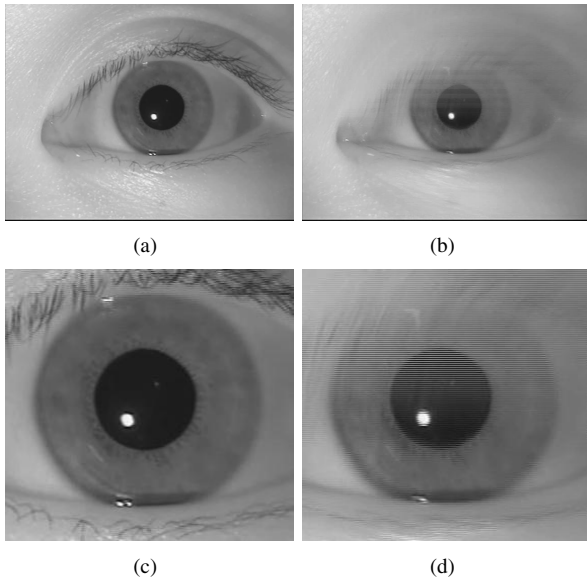
(a)   (b)

(c)   (d)

Fig. 7: Localized Motion Blur: The MSE between the current frame (a) reference frame (b) containing motion blur is 272.9. Block (c) here is the slab to be matched within a search window in frame (b) and block (d) is the matched blocked with a min MSE criterion of 696.2

equal to zero. Hence, the translational motion is confirmed and object motion without blur is successfully estimated within the reference frame.

Localized Motion Blur: An interesting point to note is that when frames with motion blur in the iris region were tested, the minimum MSE criterion using the block matching technique was found to be actually higher than the MSE between its adjacent frames. The MSE between temporal frames in Figure 7(a) and 7(b) is 272.9 but the min MSE criterion using the block matching technique is found to be higher. Figure 7(c) is the block to be matched within a given search window in 7(b), and 7(d) is the corresponding tally detected whose min MSE criterion is 696.2. This sharp rise in the MSE value can be attributed to blocks of smaller dimensions having more dissimilarity between them than frame comparisons. This also confirms the fact that this approach enables us to differentiate between those frames which are only misaligned and those which contain motion blur in the iris region. Therefore, we can conclude that a low MSE using the block matching technique would be equivalent to frame misalignment whereas a high MSE would mean the presence of motion blur. Moreover, as opposed to the motion blur detection in Section III-D where the motion blur can occur in any part of the frame, here we can validate the presence of this type of blur in the iris region.

F. Illumination & Lighting

Frames in a given video suffer from significant illumination variations. Some frames are affected globally while in other cases, only local regions of the frames are concerned. Dark frames usually have a restricted intensity range concentrated
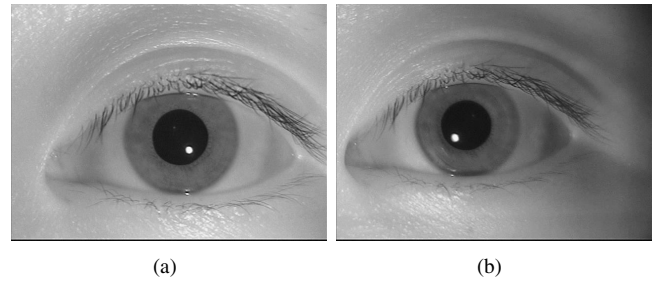

(a)   (b)

Fig. 8: Illumination Variation in Iris videos: although (a) and (b) are frames from the same video their illumination differ significantly.

to the left of the histogram while correctly illuminated frames will have a wider distribution of intensity values. The number of pixels in the upper and lower percentile region of the histogram is computed. This enables us to ascertain the type of illumination present in the images based on the number of pixels present within a given range of intensity values. Our goal here is to only detect these images rather than enhancing them. Several normalizing techniques such as Histogram Equalization or Contrast adjustment techniques could potentially be used to improve the quality of these images. Lighting also varies within a given frame. Experimentation on the given dataset reveals that on average the pupil and the iris regions make up for 12% of the pixels in a given frame and that their intensity lies between 0-0.4. In order to detect illumination variations in the iris region, the probability of pixels in that range is computed. Figure 8 shows two frames from the same iris video. The first frame is properly illuminated while the second frame has uneven illumination. Note how the RHS of the second frame is darker than the LHS. To compensate for variations in the intensity level as a result of the presence of eyelashes and eyebrows in iris frames, the averaged frame computed in Section III-A is taken as a reference point. This is shown in the cumulative distribution
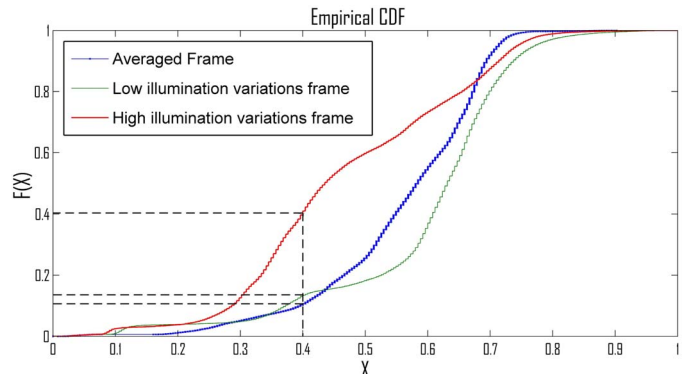

Fig. 9: Illumination Variations: A plot of the CDF of (1) the frame in Figure 8(a) with low illumination variations, (2) the frame in 8(b) with high illumination variations and (3) the averaged frame for this given video sequence

function (CDF) curves in Figure 9. It can be observed that the curve of the frame with sided illumination differs significantly from that of the averaged frame while the frame with low illumination variations closely follows that of the averaged frame. Moreover, the cumulative distribution of pixels at 0.4 is much higher in the case of the frame with sided illumination compared to both the normal and the averaged frame.

IV. Sub-optimal Videos

In this section, the selection process of the best frames in the case videos considered to be **time** and **motion** constrained is described. This usually occurs when users are non-cooperative. Figure 10(a) and Figure 10(b) are examples of the averaged frames of a Motion and Time Constrained video respectively. The absence of a global minimum in the frame contrary to that of the Adequate category occurs due to limited time and motion exposure. For example, the number of frames can vary considerably from one video to another, ranging from only 5 frames to 761 frames.

A. Time Constrained Videos

This section refers to videos with a limited number of frames i.e., fewer than 100 frames. The approach taken here is to compute the convolution of each frame with a low-pass filter. Frames with a global minimum in the central region are kept for further processing while others are discarded. The presence of a reflection in the pupil area is then used to confirm if the computed minimum does indeed correspond to the iris region. The next step is similar to the techniques implemented in Section III to verify the presence of defocus and motion blur and translational motion. Only 5 out of the 290 videos were found to belong to this category.

B. Motion Constrained Videos

As for videos in this category, the correlation coefficient was still able to pick up the correct frames after comparisons with its averaged frame. After inspecting frames in this category, we can deduce that since the number of frames is high in this instance and the majority of them contain an iris region, even though the minimum was reduced due to head movements, it was still able to separate iris frames with an iris and those with

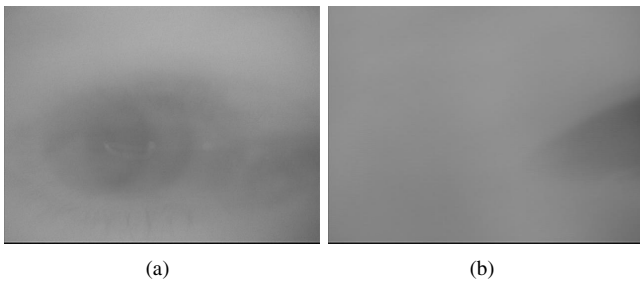

(a)   (b)

Fig. 10: Sub-optimal videos: (a) is an averaged frame of a video with significant user movement (Motion constrained) and (b) is the averaged frame of a video with limited number of frames (Time constrained).

blink and off-angle images. Furthermore, those frames which had movement in them are typically blurred and out of focus and can be captured and discarded in the further processing stages. This also acknowledges the fact that the longer the time exposure of the iris to the capturing device, the better the frame selection. 7 out of the 290 videos were found to belong to this division.

V. Discussion

In this work, the main goal was to design a highly accurate and efficient frame selection technique to enhance the performance of video based iris recognition systems The majority of previous work in this area deals with iris still images where only a few factors affecting performance is investigated [10]. On the other hand, the proposed approach takes the advantage of intra and inter frame correspondence. In [2], iris segmentation needs to be performed prior to image quality evaluation. This approach would be inefficient and impractical in the case of iris videos due high number of frames. Our proposed approach, on the other hand, using averaging is able to automatically discard frames with blink, off-angle and occlusion without performing segmentation. In [4], a fusion score is evaluated based on individual quality scores of the type of noise present in the image. However, evaluating the degree of all types of noise present in each frame of an NIR iris video is computationally expensive. Instead, a top-down approach is adopted in the case of iris videos, i.e., iris frames of poor quality are eliminated in each stage until the best ones remain as shown in the flowchart in Figure 1.

The advantage of the motion blur detection described in Section III-D is that it is very fast to compute. However, the limitation of this approach is that motion blur can occur in any region of the frame while the block matching min MSE criterion described in Section III-E ascertains that motion blur is indeed present in this iris region; however this occurs at the expense of more computational time. For the Adequate and Motion constrained category, translational motion detection is not implemented as the number of frames is high in these cases. This also ensures that the processing steps in selecting the best frames remain effective and computationally viable. However, in the case of the time constrained category where the number of frames is limited, translational motion detection is performed to recover frames which would otherwise be considered to contain motion blur. Moreover, since the number of frames is significantly reduced in this case, it remains computationally efficient.

VI. Results

The performance of a biometric system is commonly evaluated in terms of the False Accept Rate (FAR), the False Reject Rate (FRR) and the Equal Error Rate (EER). In the following experiments, for consistency and comparison purposes, the same algorithm is used to evaluate the performance of the proposed technique. All iris regions are segmented using Daugman's integro-differential operator. The iris features are
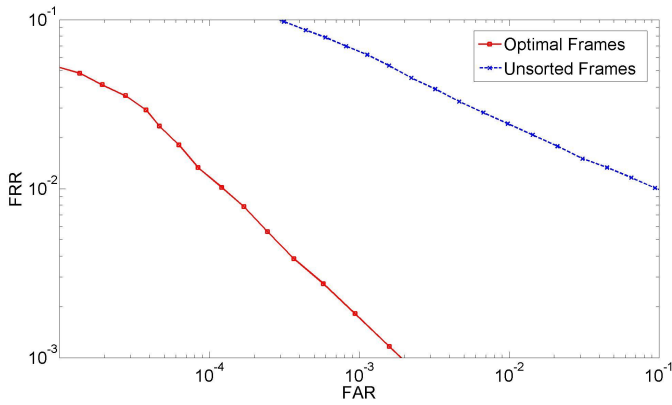
Fig. 11: ROC curves of 3000 unsorted frames in the MBGC dataset vs. 3000 optimal frames selected by the proposed method.

Method	EER(%)	d'	FRR@FAR=0.001
No frame sorting	1.88	3.90	$6.46{\times}10^{-2}$
Optimal frames	0.14	5.21	$1.73{\times}10^{-3}$

TABLE I: The selected optimal frames perform significantly better than the unsorted dataset in which no frame selection is performed.

encoded with the 1-D log Gabor wavelets and Hamming distance is used for matching [10].

NIR Iris Video: 3000 frames were randomly selected from the MBGC NIR Video dataset in which no frame selection was performed. The performance of these unsorted frames was compared with another 3000 high quality frames, i.e., frames in which severe occlusion, motion blur, out of focus and poor illumination were not detected. Figure 11 shows our results, it can be observed that optimal frames perform significantly better than the unsorted ones. Table I lists the performance metrics of unsorted vs. optimal frames. A significant drop in the EER is recorded in the case of optimal frames and the decidability index, d' increases indicating a good separation between the distribution of genuine and impostor scores. This demonstrates that the frame selection technique proposed is effective and that the performance is closely related to image quality.

NIR Iris Still vs. NIR Iris Video: The NIR Iris Still dataset consists of 1770 images of 113 subjects taken with an LG 2200 camera. Half of these images belong to the right eye and the other half are those of the left eye. The dataset consists of a wider scope of images than would usually be captured by the sensor [12]. A comparable number of optimal frames of the same subjects in the NIR Iris Video dataset are chosen using the proposed frame selection technique. The video sequences are captured with an LG 2200 iris video in this case. The performance of still images is compared with that of iris frames. The results of this comparison are shown in Figure 12 and Table II. Optimal frame selection from the iris videos achieves better performance than the iris still images. This
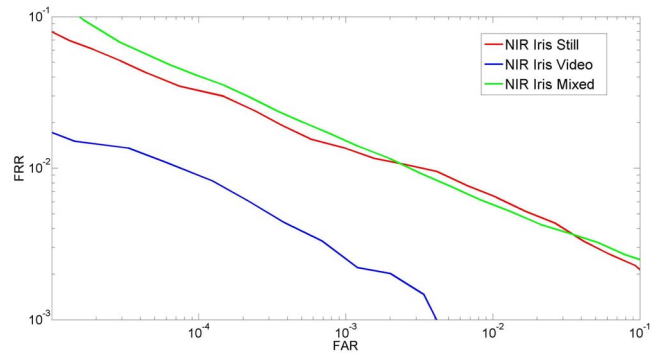

Fig. 12: ROC curve of NIR IRIS Still vs. optimal frames from the NIR Iris Video dataset. NIR Mixed dataset contains both frames and images of the same subjects.

Dataset	EER(%)	d'	FRR@FAR=0.001
NIR Iris Still	0.74	5.10	$1.34{\times}10^{-2}$
NIR Iris Mixed	0.70	4.54	$1.51{\times}10^{-2}$
NIR Iris Video	0.20	5.32	$2.52{\times}10^{-3}$

TABLE II: Optimal frames from the NIR Iris Video outperform both NIR Iris Still and NIR Iris Mixed.

highlights the reliability and accuracy of NIR Video in iris recognition systems.

NIR Iris Mixed: This dataset contains both images and frames. The number of images for each eye in the NIR Iris Still dataset is combined with an equivalent number of optimal frames of the same subjects from the NIR Iris Video dataset. Note that in this case the size of NIR Iris Mixed dataset is twice that of NIR Still dataset. As shown in Table II, there is only a slight improvement in the EER of NIR Iris Mixed compared with that of NIR Iris Still. In addition, a drop in the value of d' can also be observed. In order to determine which factor was responsible for degrading the performance, the comparisons in NIR Mixed dataset are evaluated separately. As shown in Figure 13, the FRR is the same in all three cases while the
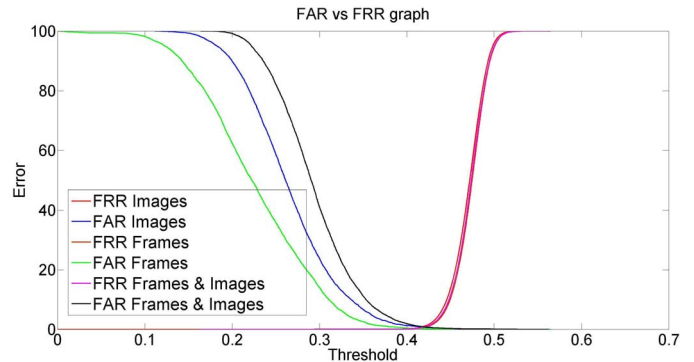

Fig. 13: Individual comparisons of the FAR and FRR of images, frames and a combination of both in the NIR Mixed dataset.

HD	Frames	Images	Frames & Images
μ_{intra}	0.22	0.26	0.30
μ_{inter}	0.47	0.47	0.47

TABLE III: Mean Hamming Distance of the intra-class and inter-class in NIR Mixed dataset.

FAR between frames and images rates poorly compared to the FAR between frames only. In Table III, μ_{inter} and μ_{intra} represent the mean Hamming Distance (HD) of the inter-class and intra-class comparisons in the NIR Iris Mixed dataset. μ_{intra} is the highest between frames and images and lowest between frames only. On the other hand, μ_{inter} is the same in all three cases. This demonstrates that interoperability between frames and images influence the HD and FAR.

Computational Time: The amount of time to process a typical adequate video is shown in Table IV. The following parameters are used in our experiments. The correlation coefficient, r is set to 0.8 for blink and off-angle detection, only frames in the upper quartile, Q_3 are selected for defocus, the min MSE criterion used for motion blur detection is set to 100.0 and finally the ratio of the number of pixels less than intensity value 0.4 in a given frame to that in the averaged frame must be less that 1.5 in order to be considered free from lighting variations. These settings can be easily varied depending on the requirements and environments.

Occlusion & Blink	Defocus $> Q_3$	Motion Blur	Illumination	Remaining Frames	Time (s)	μ_{intra}
T	F	F	F	301	4.18	0.317
T	T	F	F	43	6.67	0.251
T	T	T	F	41	7.29	0.242
T	T	T	T	33	7.63	0.223

TABLE IV: Time taken to complete the different stages of the frame selection process in an Adequate video, where T= True (Test performed) and F = False (Test not performed).

As shown in Table IV, the initial number of frames is 336 and the total amount of time for the frame selection process is found to be 7.63 seconds. In this case, 33 frames out of 336 were found to be optimal. A drop in the mean value of the HD for the intraclass comparisons can also be observed at each stage, from 0.317 to 0.223, indicating that noisy frames have been eliminated. The processing time could be further reduced by processing only the interquartile range of frames for a given video depending on the requirements.

VII. Concluding Remarks

In this paper, an optimal frame selection procedure for enhancing the performance of video based iris recognition systems is proposed. Iris videos are classified into three categories for efficient processing. Several new approaches for estimating the quality factors present in iris frames are examined. We have shown that best frames selection leads to better iris recognition performance as indicated by a drop in the EER and a rise in d' in our experiments. Our work also shows that NIR Video performs better than traditional NIR Still images and interoperability between iris frames and iris images affects the FAR. Finally, we have shown how frames of poor quality are eliminated at each stage of the proposed model in an efficient manner.

References

[1] N. Mahadeo, A. Paplinski, and S. Ray, "Model-based pupil and iris localization," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1–7.

[2] N. Kalka, J. Zuo, N. Schmid, and B. Cukic, "Estimating and fusing quality factors for iris biometric images," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 40, no. 3, pp. 509–524, 2010.

[3] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: A survey," *Comput. Vis. Image Underst.*, vol. 110, no. 2, pp. 281–307, May 2008. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2007.08.005

[4] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, "Image quality assessment for iris biometric," *Biometric Technology for Human Identification III*, vol. 6202, no. 1, p. 62020D, 2006.

[5] J. Zuo and N. Schmid, "Global and local quality measures for nir iris video," in *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, June, pp. 120–125.

[6] K. Hollingsworth, T. Peters, K. Bowyer, and P. Flynn, "Iris recognition using signal-level fusion of frames from video," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 4, pp. 837–848, 2009.

[7] R. Jillela, A. Ross, and K. Bowyer, "Information fusion in low-resolution iris videos using principal components transform," in *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, 2011, pp. 262–269.

[8] Y. Lee, P. J. Phillips, and R. J. Micheals, "An automated video-based system for iris recognition," in *Proceedings of the Third International Conference on Advances in Biometrics*, ser. ICB '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 1160–1169. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-01793-3_117

[9] Z. Zhou, Y. Du, and C. Belcher, "Transforming traditional iris recognition systems to work in nonideal situations," *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 8, pp. 3203–3213, Aug.

[10] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 21–30, 2002.

[11] B. Kang and K. Park, "A study on iris image restoration," in *Audio- and Video-Based Biometric Person Authentication*, ser. Lecture Notes in Computer Science, T. Kanade, A. Jain, and N. Ratha, Eds. Springer Berlin Heidelberg, 2005, vol. 3546, pp. 31–40. [Online]. Available: http://dx.doi.org/10.1007/11527923_4

[12] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, Iii, and S. Weimer, "Overview of the multiple biometrics grand challenge," in *Proceedings of the Third International Conference on Advances in Biometrics*, ser. ICB '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 705–714. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-01793-3_72

[13] J. L. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia, "Diatom autofocusing in brightfield microscopy: a comparative study," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 3, pp. 314–317 vol.3.

[14] K. Thyagarajan, *Still Image and Video Compression with MATLAB*. Wiley, 2011. [Online]. Available: http://books.google.com.au/books?id=adv71MkRIWYC