Improving Video-based Iris Recognition Via Local Quality Weighted Super Resolution

Nadia Othman[1], Nesma Houmani[2] and Bernadette Dorizzi[1]
[1]*Institut Mines-Télécom, Télécom SudParis, 9 rue Charles Fourier, Evry, France*
[2]*Laboratoire SIGMA, ESPCI-ParisTech, 10 rue Vauquelin, Paris, France*

Keywords: Iris Recognition, Video, Quality, Super Resolution, Fusion of Images.

Abstract: In this paper we address the problem of iris recognition at a distance and on the move. We introduce two novel quality measures, one computed Globally (GQ) and the other Locally (LQ), for fusing at the pixel level the frames (after a bilinear interpolation step) extracted from the video of a given person. These measures derive from a local GMM probabilistic characterization of good quality iris texture. Experiments performed on the MBGC portal database show a superiority of our approach compared to score-based or average image-based fusion methods. Moreover, we show that the LQ-based fusion outperforms the GQ-based fusion with a relative improvement of 4.79% at the Equal Error Rate functioning point.

1 INTRODUCTION

The excellent performance of biometric systems based on the iris are obtained by controlling the quality of the images captured by the sensors, by imposing certain constraints on the users, such as standing at a fixed distance from the camera and looking directly at it, and by using algorithmic measurements of the image quality (contrast, illumination, textural richness, etc.).

However, when working with moving subjects, as in the context of surveillance video or portal scenarios for border crossing, many of these constraints become impossible to impose. An "iris on the move" (IOM) person recognition system was evaluated by the NIST by organizing the Multiple Biometric Grand Challenge (MBGC, 2009). The image of the iris is acquired using a static camera as the person is walking toward the portal. A sequence of images of the person's face is acquired, which normally contain the areas of the eyes.

The results of the MBGC show degradation in performance of iris systems in comparison to the IREX III evaluation based on databases acquired in static mode. With a 1% false acceptance rate (FAR), the algorithm that performed best in both competitions obtains 92% of correct verification on the MBGC database, as compared to 98.3% on the IREX III database.

Indeed, acquisition from a distance causes a loss in quality of the resulting images, showing a lack of resolution, often presenting blur and low contrast between the boundaries of the different parts of the iris.

One way to try to circumvent this bad situation is to use some redundancy arising from the availability of several images of the same eye in the recorded video sequence. A first approach consists in fusing the scores coming from the frame by frame matching (1 to 1) by some operators like the mean or the min. This has been shown to be efficient but at the price of a high computational cost (Hollingsworth et al., 2009). Another direction is to fuse the images at the pixel level, exploiting this way the redundancy of the iris texture at an early stage and to perform the feature extraction and matching steps on the resulting fused images. At this point, the remaining question is how to perform this fusion stage so that the performance can be improved compared to 1 to 1 or score fusion schemes.

At our knowledge, few authors have considered the problem of fusing images of low quality in iris videos for improving recognition performance. The first paper is that of Fahmy (2007) who proposed a super resolution technique based on an auto-regressive signature model for obtaining high resolution images from successive low resolution ones. He shows that the resulting images are

Othman N., Houmani N. and Dorizzi B..
Improving Video-based Iris Recognition Via Local Quality Weighted Super Resolution.
DOI: 10.5220/0004342306230629
In *Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods* (BTSA-2013), pages 623-629
ISBN: 978-989-8565-41-9

valuable only if the initial low-resolution images are blur-free and focused, stressing already the bad influence of low quality images in the fusion. In (Hollingsworth et al., 2009), authors proposed to perform a simple averaging of the normalized iris images extracted from the video for matching NIR videos against NIR videos from the MBGC database. When compared to a fusion of scores, the results are similar but with a reduced complexity. In the same spirit, Nguyen et al., (2010; 2011b) proposed to fuse different images of the video at a pixel level after an interpolation of the images. They use a quality factor in their fusion scheme, which allows giving less importance to images of bad quality. The interpolation step is shown very efficient as well as the quality weighting for improving recognition performance. Note that they considered a protocol similar to MBGC, where they compare a video to a high quality still image. More recent papers (Nguyen et al., 2011a); (Jillela et al., 2011) explored the fusion in the feature domain using PCA or PCT but not on the same MBGC protocol as they usually degrade artificially the image resolution in their assessment stage.

In our work, like in (Nguyen et al., 2011b), we propose to fuse the different frames of the video at the pixel level, after an interpolation stage which allows increasing the size of the resulting image by a factor of 2. Contrary to (Nguyen et al., 2011b), we do not follow the MBGC protocol which compares a video to a still high quality image reference but we consider in our work, a video against video scenario, more adapted to the re-identification context, meaning that we will use several frames in both low quality videos to address the person recognition hypothesis.

The above literature review dealing with super resolution in the iris on the move context has stressed the importance of choosing adequately the images involved in the fusion process. Indeed, integration of low quality images leads to a decrease in performance producing a rather counterproductive effect.

In this work, we will therefore concentrate our efforts in the proposition of a novel way of measuring and integrating quality measures in the image fusion scheme. More precisely our first contribution is the proposition of a global quality measure for normalized iris images as defined in (Cremer et al., 2012) as a weighting factor in the same way as proposed in (Nguyen et al., 2011b). The interest of our measure compared to (Nguyen et al., 2011b) is its simplicity and the fact that its computation does not require to identify in advance the type of degradations that can occur. Indeed our measure is based on a local GMM-based characterization of the iris texture. Bad quality normalized iris images are therefore images containing a large part of non-textured zones, resulting from segmentation errors or blur.

Taking benefit of this local measure, we propose as a second novel contribution to perform a local weighting in the image fusion scheme, allowing this way to take into account the fact that degradations can be different in different parts of the iris image. This means that regions free from occlusions will contribute more in the reconstruction of the fused image than regions with artifacts such as eyelid or eyelash occlusion and specular reflection. Thus, the quality of the reconstructed image will be better and we expect this scheme to lead to a significant improvement in the recognition performance.

This paper is organized as follows: in Section 2 we describe our approach for Local and Global quality based super resolution and in Section 3 we present the comparative experiments that we performed on the MBGC database. Finally, conclusions are given in Section 4.

2 LOCAL AND GLOBAL QUALITY-BASED SUPER RESOLUTION

In this Section, we will first briefly describe the different modules of a video-based iris recognition system. We will also recall the definition of the local and global quality measure that we will use on the normalized images. This concept has been described in details in (Cremer et al., 2012); (Krichen et al., 2007). We will explain how we have adapted this measure to the context of iris images resulting from low quality videos. We also describe the super-resolution process allowing interpolation and fusion of the frames of the video. Finally, we will summarize the global architecture of the system that we propose for person recognition from moving person's videos using these local and global quality measures.

2.1 General Structure of Our Video-based Iris Verification System

For building an iris recognition system starting from a video, several steps have to be performed. The first need is the detection and tracking of the eyes in the sequence, generally guided by the presence of spots that are located around the eyes. Once this stage has been completed, very poor quality images in the sequence are discarded and on the remaining frames, the usual segmentation and normalization steps of the iris zone must be performed.

In this work, we use the MBGC database. One of the difficulties present in this database lies in the fact that light spots, which can cause errors when looking for the boundaries of the iris, often occlude the boundary between the iris and the pupil. For this reason, we perform a manual segmentation of the iris boundaries, which provides normalization circles.

We then use the open source iris recognition system OSIRISv2, inspired by Daugman's approach (Daugman, 2002), which was developed in the framework of the BioSecure project (BioSecure, 2007). More precisely, as previously said, we do not use the segmentation stage of OSIRISv2 but only the normalization, feature extraction and matching steps. For finding the occlusion masks, we use an adaptive filter similar to the one proposed in (Sutra et al., 2012) but adapted to the case of images extracted from a video sequence.

2.2 Local Quality Measure

As in (Krichen et al., 2007), we use a Gaussian Mixture Model (GMM) to give a probabilistic measure of the quality of local regions of the iris. In this work, the GMM is learned on small images extracted from the MBGC database showing a good quality texture free from occlusions. So, this model will give a low probability on the noisy regions, which result from blur or artifacts as shown in (Cremer et al., 2012). The interest of this approach is that there is no need to recognize in advance the type of noise present in the images such as eyelid or eyelash occlusion, specular reflection and blur.

We trained the GMM with 3 Gaussians on 95 sub-images free from occlusions, selected manually from 30 normalized images taken randomly from MBGC database. In the same way as in (Cremer et al., 2012), the model is based on four local observations grouped in the input vector x_i: the intensity of the pixel i, the local mean, the local variance and the local contrast measured in a 5x5 neighbourhood of the pixel i. The quality measure associated to a sub-image (w) of an image is given by the formula:

$$Q_{local}(w) = exp^{-\frac{1}{d}\sum_{i=1}^{d}|\log(p(x_i/\lambda)) - \bar{\alpha}|} \qquad (1)$$

Where d is the size of the sub-image (w), x_i is the input vector of our GMM, $p(x_i/\lambda)$ is the likelihood given by the GMM λ to the input vector x_i, and $\bar{\alpha}$ is the mean log-likelihood on the training set. We use a negative exponential to obtain a value between 0 and 1. The closest Q value will be to 1, the highest are the chances that the sub-image w is of good quality, namely free from occlusion and highly textured.

2.3 Global Quality Measure

The local measure presented in Section 2.2 can also be employed for defining a global measure of the quality of the entire image. To this end, we divide the normalized image (of size 64x512) in overlapping sub-images of size 8x16 and we average the probabilities given by the local GMM of each sub-image as follows:

$$Q_{global} = \frac{1}{N}\sum_{n} Q_{local}(w_n) \qquad (2)$$

Where N is the number of sub-images and $Q_{local}(w_n)$ is the GMM local quality of the nth sub-image.

2.4 Super Resolution Implementation

MBGC's images suffer from poor resolution, which degrades significantly iris recognition performance. Super resolution (SR) approaches can remedy to this problem by generating high-resolution images from low-resolution ones. Among the various manners to implement SR schemes, we chose in this work a simple version similar to that exploited in (Nguyen et al., 2010), resulting into a double resolution image using the bilinear interpolation.

After interpolating each normalized image of the sequence, a step of registration is generally needed before pixel's fusion to ensure that those pixels are correctly aligned with each other in the sequence. But for MBGC videos, authors are divided on the fact that the images of the sequence need to be registered. We tried performing some registration by identifying the shift value that maximized the phase correlation between the pixel values and we noticed that registration didn't produce any better recognition performance. Indeed, the process of normalization already performs a scaling of the iris zone, allowing an alignment of the pixels, which is sufficient for the present implementation of super resolution.

This set of normalized interpolated images is then fused to obtain one high-resolution image. We introduce some quality measures in this fusion process. More precisely, as done in (Nguyen et al., 2010), we weight the value of each pixel of each image by the same factor, namely the Global Quality (GQ) (defined in Section 2.3) of the corresponding image. We also propose a novel scheme using our Local Quality (LQ) measure (defined in Section 2.2). In this latter case, we compute the quality measures of all the sub-images as defined in Section 2.3 and we generate a matrix of the same size as the normalized image which contains the values of the quality of each sub-images. This matrix is then bilinearly interpolated. Finally, we weight the value of each pixel of each interpolated image by its corresponding value in the interpolated quality matrix. Figure 1 illustrates this LQ-based fusion process which is more detailed in Section 2.5.

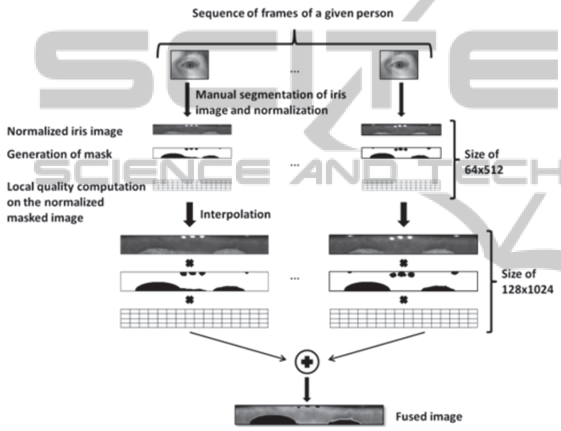

Figure 1: Fusion process of the proposed local quality-based method.

2.5 Architecture of the Local Quality-based System

Figure 2 presents the general architecture of our LQ-based system. The main steps of such system are described as follows:

- Discard very low quality (highly blurred) frames of the sequence using wavelet's transform.

Then, for each frame:

- Detect and extract the periocular zone,
- Segment manually the iris using two non-concentric circles approximation for the pupillary and limbic boundaries,
- Normalize the segmented iris zone with Daugman's rubber sheet technique,
- Generate masks and measure the local quality on the normalized and masked images, using the GMM already learned,
- Interpolate the normalized images and their corresponding masks and local quality matrices to a double resolution using the bilinear interpolation.

Finally, for all frames, generate the fused image as follows:

$$I_{fused} = \frac{\sum_{i=1}^{F} I^i(x,y) * M^i(x,y) * Q^i(\mathrm{w})}{\sum_{i=1}^{F} M^i(x,y) * Q^i(w)} \quad (3)$$

Where F is the total number of frames, $I^i(x,y)$ and $M^i(x,y)$ are the values of the pixel in the position (x,y) of, respectively, the ith interpolated normalized image and mask. $Q^i(w)$ is the local quality of the sub-image (w) to which the pixel (x,y) belongs.

The last steps of the recognition process namely feature extraction and matching (as recalled previously in Section 2.1) are performed on the fused reconstructed image. Note that from one video of *F* frames, we get only one image performing this way an important and efficient compression of the information.

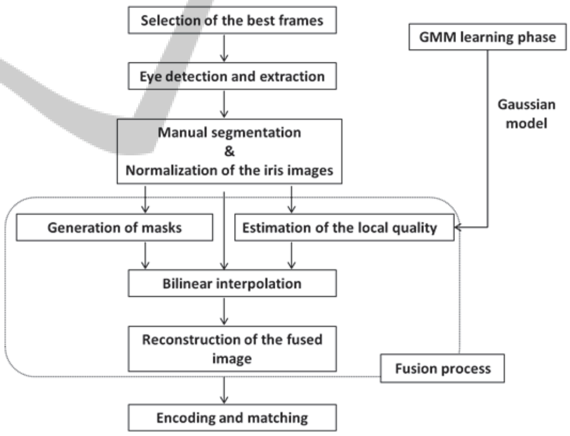

Figure 2: Diagram of the Local Quality-based system for video-based iris recognition.

3 EXPERIMENTS AND RESULTS

3.1 Database and Protocols

The proposed method has been evaluated on the portal dataset composed of Near Infra-Red (NIR) faces videos used during the Multiple Biometric Grand Challenge organized by the National Institute of Standards and Technology (MBGC, 2009). This dataset called MBGC was acquired by capturing facial videos of 129 subjects walking through a portal located at 3 meters from a NIR camera.

Although the resolution of the frames in the video is 2048x2048, the number of pixels across the iris is about 120, which is below the minimum of 140 pixels considered as the minimum to ensure a good level of performance. The images suffer not only from low resolution but also from motion blur, occlusion, specular reflection and high variation of illumination between the frames. Examples of poor quality images can be found in Figure 3.

Due to the important variation of illumination that can be observed between the frames across one sequence, we manually discard darker ones as done in other work. After that, blurred frames from the sequence were removed by using wavelet's transformation. After all these pre-processing, the database is composed of 108 subjects and each one possesses 2 sequences with at least 4 frames per sequence.

We didn't follow the protocols specified in MBGC. Indeed, we didn't compare still images to videos as in (Nguyen et al., 2011b) but NIR videos to NIR videos like in (Hollingsworth et al., 2009). For each person, we use the first sequence as a target and the second one as a query.

3.2 Experiments and Results

The proposed approach is compared to other fusion score methods such as Multi-Gallery Simple-Probe (MGSP), Multi-Gallery Multi-Probe (MGMP) and also to fusion signal methods as simple averaging of images and weighted super resolution.

3.2.1 Fusion at the Score Level

- Matching 1 to 1: all the frames in the video of a person are considered as independent images and used for performing inter-class and intra-class comparisons. This system was used as a baseline system to compare the other methods.
- Matching N to 1, Multi-Gallery Simple-Probe: in this case, the different images in the video are considered dependent as they represent the same person. If the number of samples in the gallery and the probe are respectively N and 1 per person, we get N Hamming distance scores which can be fused by making a simple average (Ma et al., 2004) or the minimum of all the scores (Krichen et al., 2005).
- Matching N to M, Multi-Gallery Multi-Probe: in this case, we consider M images in the probe and N images in the gallery. We thus get N*M scores per person and combine them by taking the average or the minimum.

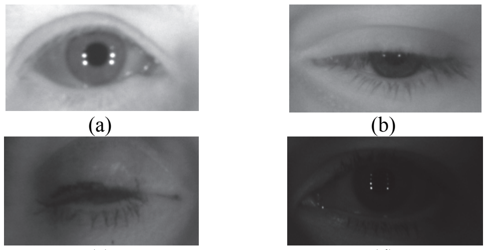

Figure 3: Examples of bad quality images: a) out of focus, b) eyelid and eyelashes occlusions, c) closed eye, d) dark contrast.

The performance of these score fusion schemes are shown in Table 1.

Table 1: Equal Error Rate (EER) of the score's fusion methods.

Methods	**EER (in %)**	
Matching 1 to 1 (baseline)	14.32	
	Minimum	Average
Matching 1 to *N* (MGSP)	9.30	10.27
Matching M to N (MGMP)	4.66	5.65

As shown in Table 1, the best score's fusion scheme reduces the Equal Error Rate (EER) from 14.32% to 4.66%. This indicates that recognition performance can be further improved by the redundancy brought by the video. However, the corresponding matching time increases considerably when the recognition score is calculated for N*M matchings.

3.2.2 Fusion at the Signal Level

- Without quality: At first, the fusion of images is done without using quality measure. For each sequence, we create a single image by averaging the pixels intensities of the different frames of such a sequence. We experimented two cases: with and without interpolated images. The EER of the two methods is reported in Table 2.

Table 2: Equal Error Rate (EER) of the image's fusion methods without using quality.

Strategy of fusion	**EER (in %)**
Simple average of normalized iris	4.90
Simple average of interpolated normalized iris (SR)	3.66

Table 2 shows that the fusion method based on the interpolation of images before averaging the pixel intensities outperforms the simple average method with a relative improvement of 25.30% at the EER functioning point. This result is coherent with Nguyen's results which states that super resolution (SR) greatly improves recognition performance (Nguyen et al., 2010).

By observing Table 1 and Table 2, we see that the MPMG-min method is slightly better than the simple average (4.66% vs 4.9%). These results are coherent with those obtained by Hollingsworth et al (2009). However, as explained in their work, the matching time and memory requirements are much lower for image's fusion than score's fusion.

- With quality (global and local): Given the considerable improvement brought by the interpolation, we decided to perform our further experiments only on SR images. We introduce in the fusion the global quality (GQ) and local quality (LQ) fusion schemes as explained in Section 2.4. The Equal Error Rate (EER) of all methods is shown in Table 3 and the DET-curves of these methods are shown in Figure 4.

As shown in Table 3, introducing our global quality criterion in the fusion gives a high relative recognition improvement (25.95% at the EER). Our method is in agreement with Nguyen's result (Nguyen et al., 2011b) who obtains an improvement of 11.5% by introducing his quality measure (but with another evaluation protocol). Compared to his method, our quality is simpler to implement. Indeed, the metric employed by Nguyen et al. (2011b) to estimate the quality of a given frame includes four independent factors: focus, off-angle, illumination variation and motion blur. After calculating individually each of these quality scores, a single score is obtained with the Dempster-Shafer theory. Our quality measure has the advantage of not requiring extra strategy of combinations neither knowing in advance the possible nature of the degradation.

Table 3: Equal Error Rate (EER) of the image's fusion methods with and without quality measures.

Strategy of fusion	EER (in%)
Without quality	3.66
With global quality	2.71
With local quality	2.58

By incorporating our GQ measure in the fusion process, the contribution of each frame in the fused image will be correlated to its quality, this way more weight is given to the high quality images.

Table 3 also shows that LQ-based fusion method outperforms the GQ-based fusion method with a relative improvement of 4.79% at the EER. This is due to the fact that the quality in an iris image is not globally identical: indeed, due for example to motion blur, a region in an iris image could be more textured than another one. Moreover, our LQ measure can detect eventual errors of masks and assign them a low value. The LQ-based fusion scheme allows therefore a more accurate weighting of the pixels in the fusion scheme than the GQ method.

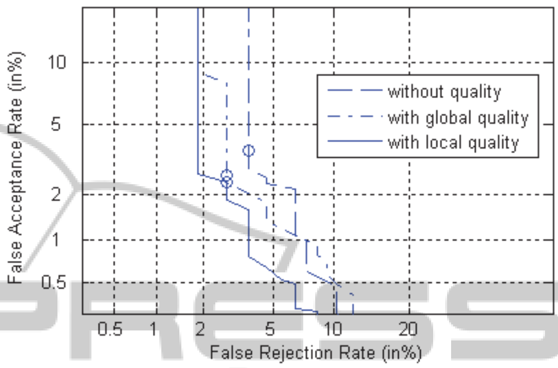

Figure 4: DET-curves of the three image's fusion approaches.

4 CONCLUSIONS

In this paper, we have proposed two novel contributions for implementing image fusion of frames extracted from videos of moving persons with the aim of improving the performance in iris recognition. Our main novelty is the introduction in the fusion scheme, at the pixel level, of a local quality (LQ) measure relying on a GMM estimation of the distribution of a clean iris texture. This LQ measure can also be used for giving a global quality (GQ) measure of the normalized iris image. We have shown on the MBGC database that the LQ-based fusion allows a high improvement in performance compared to other fusion schemes (at the score or image level) or to our GQ-based fusion.

The present work corresponds to a first step towards the production of a global and automatic system able to process in real time, videos acquired in an optical gate. In fact we have so far only validated our approach using some manual interventions for the first steps of the process (choice of adequate images and iris segmentation) and new modules would be necessary for building an automatic system. More precisely, we have made a manual selection of the very low quality images (as done by most authors in the field) but this could be performed thanks to a simple global quality measure. An automatic segmentation procedure can replace the manual one but, due to the low quality of MBGC frames, we expect that it will produce a large number of errors (as assessed by the degradation of performance observed in the MBGC competition). However our intuition is that our local quality measure should be able to detect those errors and that our system will therefore be able to discard those bad-segmented pixels from the fusion procedure. If this is the case our fusion procedure should not suffer too much from segmentation errors. Our future works will tend to validate this hypothesis using a bigger database with more videos per person.

REFERENCES

BioSecure Project (2007). http://biosecure.it-sudparis.eu.

Cremer, S., Dorizzi, B., Garcia-Salicetti S. and Lempiérière, N., (2012). How a local quality measure can help improving iris recognition». In *Proceedings of the International Conference of the Biometrics Special Interest Group.*

Daugman, J., (2007). How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, p. 21–30.

Fahmy, G., (2007). Super-resolution construction of IRIS images from a visual low resolution face video. *In Proceedings of the International Symposium on Signal Processing and Its Applications*.

Hollingsworth, K., Peters, T., Bowyer, K. W. and Flynn, P. J., (2009). Iris recognition using signal-level fusion of frames from video. *IEEE Transactions on Information Forensics and Security*, vol. 4, n^{o}. 4, p. 837–848.

Jillela, R., Ross, A. and Flynn, P. J., (2011). Information fusion in low-resolution iris videos using Principal Components Transform. In *Proceedings of the 2011 IEEE Workshop on Applications of Computer Vision* p. 262–269.

Krichen, E., Garcia-Salicetti, S., and Dorizzi, B., (2007). A new probabilistic Iris Quality Measure for comprehensive noise detection. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems*, p. 1-6.

Krichen, E., Allano, L., Garcia-Salicetti, S., and Dorizzi, B., (2005). Specific Texture Analysis for Iris Recognition. In *Audio- and Video-Based Biometric Person Authentication*, Springer Berlin Heidelberg, vol. 3546, p. 23-30.

Ma, L., Tan, T., Wang, Y. and Zhang, D., (2004). Efficient Iris Recognition by Characterizing Key Local Variations. *IEEE Transactions on Image Processing*, vol. 13, p. 739–750.

MBGC, (2009). MBGC Portal Challenge Version 2 Preliminary Results. *National Institute of Standards and Technology – MGBC 3rd Workshop.* http://www.nist.gov/itl/iad/ig/mbgc-presentations.cfm.

Nguyen, K., Fookes, C. B., Sridharan, S. and Denman, S., (2010). Focus-score weighted super-resolution for uncooperative iris recognition at a distance and on the move. In *Proceedings of the International Conference of Image and Vision Computing*.

Nguyen, K., Fookes, C. B., Sridharan, S. and Denman, S., (2011a). Feature-domain super-resolution for IRIS recognition. In *Proceedings of the International Conference on Image Processing.*

Nguyen, K., Fookes, C. B., Sridharan, S. and Denman, S., (2011b). Quality-Driven Super-Resolution for Less Constrained Iris Recognition at a Distance and on the Move. *IEEE Transactions on Information Forensics and Security*, vol. 6, n^{o}. 4, p. 1248-1258.

Sutra, G., Garcia-Salicetti, S. and Dorizzi, B., (2012). The Viterbi algorithm at different resolutions for enhanced iris segmentation. In *Proceedings of the International Conference on Biometrics*, p. 310-316.