

# A Study on Quality-adjusted Impact of Time Lapse on Iris Recognition

Nadezhda Sazonova<sup>a</sup>, Fang Hua<sup>b</sup>, Xuan Liu<sup>c</sup>, Jeremiah Remus<sup>b</sup>, Arun Ross<sup>d</sup>, Lawrence Hornak<sup>d</sup>,  
Stephanie Schuckers<sup>b</sup>

<sup>a</sup>ECE Department, University of Alabama, Tuscaloosa, AL, 35487; <sup>b</sup>ECE Department, Clarkson University, Potsdam, NY, 13699; <sup>c</sup>ECE Department, University of Florida, Gainesville, FL, 32611;

<sup>d</sup>Lane Department of CSEE, West Virginia University, Morgantown, WV, 26506

## ABSTRACT

Although human iris pattern is widely accepted as a stable biometric feature, recent research has found some evidences on the aging effect of iris system. In order to investigate changes in iris recognition performance due to the elapsed time between probe and gallery iris images, we examine the effect of elapsed time on iris recognition utilizing 7,628 iris images from 46 subjects with an average of ten visits acquired over two years from a legacy database at Clarkson University. Taken into consideration the impact of quality factors such as local contrast, illumination, blur and noise on iris recognition performance, regression models are built with and without quality metrics to evaluate the degradation of iris recognition performance based on time lapse factors. Our experimental results demonstrate the decrease of iris recognition performance along with increased elapsed time based on two iris recognition system (the modified Masek algorithm and a commercial software VeriEye SDK). These results also reveal the significance of quality factors in iris recognition regression indicating the variability in match scores. According to the regression analysis, our study in this paper helps provide the quantified decrease on match scores with increased elapsed time, which indicates the possibility to implement the prediction scheme for iris recognition performance based on learning of impact on time lapse factors.

**Keywords:** Iris recognition performance, time lapse, statistical analysis, robust regression, iris quality measurement

## 1. INTRODUCTION

The stability of the human iris pattern over a long period of time is a broadly accepted assumption in the biometric scientific community. Although it has been shown that iris recognition systems can match the iris images of an individual taken many years apart [1][2], the overall sensitivity of match scores to elapsed time has not been studied until recently. Gonzalez et al. [3] found the variability on a shorter time scale while data was collected during sessions that were one to four weeks apart, with up to four sessions for some subjects. They showed false rejection rates that increased as a function of the time between sessions at which the images were collected. Then Fenker et al. [4] presented a statistically significant change in the average Hamming distance between iris images taken four years apart even considered possible effects from camera degradation. However, Baker et al's study was based on a relatively small data set consisting of 13 subjects (26 irises) and a total of 1809 images. These two studies were trying to demonstrate the degradation of iris performances along with elapsing time, but ignored other possible factors for the impact. Fenker and Bowyer [5] investigated the effect of two factors on the iris template aging: iris dilation and the presence of contact lenses. They found some evidences that both the pupil dilation and the contact lenses could be the significant confounding factors for measuring iris template aging.

This conclusion enlightened us that even use the same acquisition equipment over a period of time, iris recognition performance still could be affected by other factors. Many researches have proved that quality of images could affect the recognition performance in biometric systems [6][7][8][9]. Hence, we put our efforts in this paper on the analysis of impact of elapsed time on iris recognition utilizes a large number of iris images taken over two years with adjustment for quality factors from iris images. Different quality metrics, such as occlusion, local contrast, illumination and sharpness are incorporated into the statistical model which relates iris recognition performance as a function of time lapse.

The rest of this paper is organized as follows. Section 2 briefly describes four quality measures and presents the statistical modeling with and without quality factors followed with a summary of two iris matching algorithms. Section 3 introduces the database and presents experimental results with the analysis. At last, a conclusion is summarized in Section 4.

## 2. METHODS

### 2.1 Quality measurements

Here, we introduce four factors for iris quality measurement: occlusion, local contrast, illumination, and Sharpness.

Occlusion is the most seen factors on iris images because it usually generated by eyelashes, specula reflections or eye lids, etc. A noise mask is designed to detect the occlusion on each iris image to avoid non-iris pattern. The quality metric for occlusion is computed as the percentage of noisy pixels in the normalized iris image given by the noise mask (Eq 1). Thereupon, other quality factors are evaluated based on the un-noise areas of iris region.

$$OC = \frac{n_{noisy}}{n \times m} = \frac{n \times m - N}{n \times m} \quad (1)$$

Since iris images require reasonably clear textural patterns to be used for matching, the measure of the contrast should take into account local variations within an image. Root Mean Square (RMS) has been presented in [10] to measure the contrast of an image, calculating the standard deviation of the pixel intensities. Here, we introduce our version of RMS to measure the local contrast in iris images (Eq 2):

$$LC = \sqrt{\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m \delta(I_{ij} - M_{ij})^2} \quad (2)$$

where  $n$  and  $m$  are the vertical and horizontal dimensions of a normalized iris image.  $N$  is the number of non-noisy pixels.  $I_{ij}$  is the intensity of pixel  $(i, j)$ ,  $M_{ij}$  is the median of the intensities of pixels within a  $10 \times 10$  window centered at  $(i, j)$ , and  $\delta$  is noise-pixel identifier which is equal to zero if pixel  $(i, j)$  is found to be noisy and otherwise is equal to 1.

The illumination condition can yield a big impact on iris images with different brightness on iris images. Hence, the average grayscale intensity of the entire image is computed simply as the illumination metric in Eq 3:

$$IL = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m I_{ij} \delta \quad (3)$$

where  $n$ ,  $m$  are the vertical and horizontal dimensions of a normalized iris image.  $N$  is the number of non-noisy pixels.

According to Wan et al. [11], the LoG-based sharpness metric is able to detect out-of-focus as well as motion blur. Hence, the sharpness of an image is computed as the mean value of the non-noisy pixels in the LoG-filtered image:

$$SH = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m I_{ij}^{LoG} \delta \quad (4)$$

$$LoG(x, y) = -\frac{1}{\pi\sigma^4} \left[ 1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (5)$$

where LoG present the 2D Laplacian of Gaussian Operator (Eq 5). Two parameters for controlling this 2-D LoG filter are window size ( $hs$ ) and standard deviation ( $\sigma$ ). They are implemented in the algorithm as 5 and 2.0, respectively.

### 2.2 Statistical regression models

In order to understand the relationship between the match scores and time lapse, we conduct the robust regression analysis for estimating the tendency of performance degradation in match scores along with time lapse. Robust regression analysis is less sensitive to outliers and can utilize majority data, which is a reasonable choice for biometric system. The iteratively reweighted least-squares (IRLS) method is used to find an M-estimator for the robust regression model. The basic robust regression model is presented in Eq 6. The quality-adjusted robust regression model is shown in Eq 7:

$$HD = \alpha_0 + \alpha_1 t + \varepsilon \quad (6)$$

$$HD_q = \beta_0 + \beta_1 t + \beta_{21} q_{11} + \beta_{22} q_{12} + \dots + \varepsilon \quad (7)$$

where  $t$  is the time lapse as the predictor variable. In Eq 6,  $HD$  is match scores from Hamming distance which is the

dependent variable to time only;  $\alpha_0$  is the intercept,  $\alpha_1$  is the regression coefficient. In Eq 7,  $HD_q$  is match scores which are dependent to time and four quality metrics;  $\beta_0$  is the intercept,  $\beta_1$  is the coefficient for  $t$ , and  $q_{i1}$  and  $q_{i2}$  are  $i^{\text{th}}$  quality metric values for images 1 and 2, respectively.

### 2.3 Iris recognition algorithms

Two iris recognition algorithms are using for generating match scores in this study: modified Libor Masek method[12], and a commercial software VeriEye SDK from Neurotechnology Inc. We apply our modified version of the Libor Masek algorithm with Gabor encoding to generate the match scores for iris images. A better segmentation approach for iris images based on the relative entropy of grayscale values across the circular boundary is integrated in the Libor Masek method. Matching results for the modified Libor Masek algorithm is performed using the traditional Hamming distance metrics while match scores from VeriEye SDK are similarity measures between matched images.

## 3. RESULTS AND DISCUSSION

### 3.1 Database

To evaluate the iris performance with time lapse, a subset from the Multimodal Biometric Dataset [13] from Clarkson University Bio-Signal Analysis Laboratory is considered to use in this work. Iris images from 244 subjects are collected by the same camera (OKI IRISPASS-h) from year 2005 to year 2007 resulting in a total of 7628 images. Figure 2 shows the distribution of subjects across maximal time lapses for each subject between the first and the last visit. Out of 244 subjects, 46 subjects have images with time lapses over 1 year; 15 subjects have images with time lapses over 2 years. Here, we chose 46 subjects who are collected iris images with at least one year between first visit and last visit. Figure 1 shows some example iris images from the dataset demonstrating different quality factors as occlusion, illumination, sharpness and local contrast variations across images. Figure 3 and Figure 4 show the maximum time lapse and number of visits per selected subject respectively.

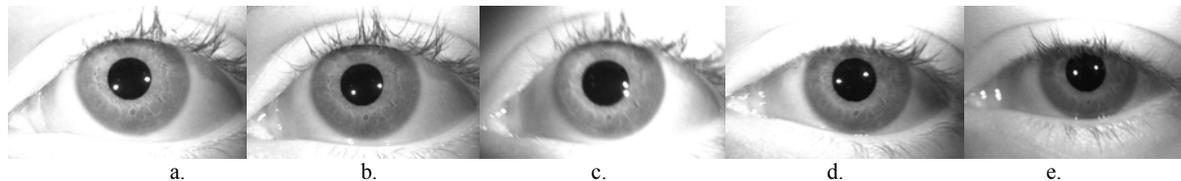


Figure 1, Example iris images from the same subject in the database. a. is the best image; b. has different local contrast than a.; c. is a blurry image; d. has a lower illumination compare to a; e. has more occlusion than a.

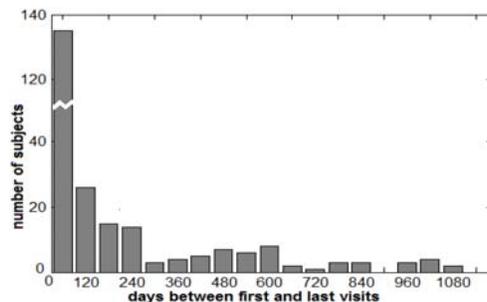


Figure 2, Distribution of all 244 subjects in the database with respect to the time elapsed between the first and the last visits.

### 3.2 Regression of quality factors

Despite the reason for degradation of iris performance along with elapsed time is unclear, it is still important to evaluate the tendency of iris qualities over time. Besides the average quality of each month, the regression of quality metrics of occlusion, local contrast, illumination and blur for all 7628 images from 244 subjects are measured and presented in the chronological order.

Figure 5 shows quality measures as a function of time in months along with the fitted robust regression lines compared with medians. Local contrast, illumination and occlusion show the slight negative slope in their regression lines

indicating a slight decrease in these quality metrics over time. Conversely, the blur metric exhibits some degree of increase over time which presents degradation on sharpness.

Despite some evidences toward the decrease in overall quality of iris images over time using four metrics described above, the actual time trend of the quality of iris images is hard to evaluate due to subjective selection of quality metrics as well as the relatively low changes in these metrics over time. In the next section, the four quality measures are incorporated into the model to study the change in match scores with increased time difference between images. Further, we consider the significance of this linear fit for some of the quality factors.

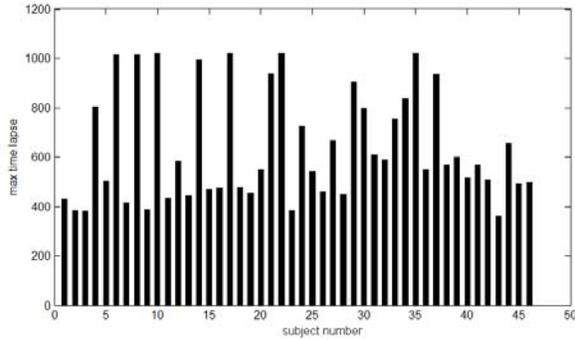


Figure 3, Maximum time lapse in match scores for selected 46 subjects.

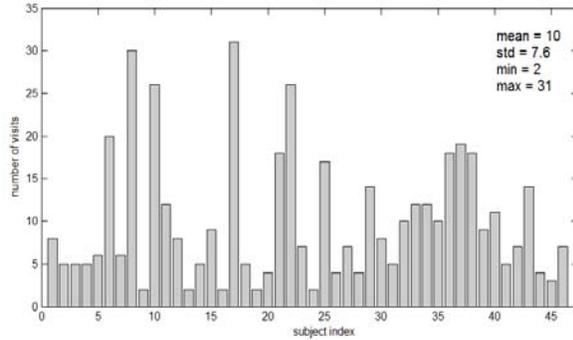


Figure 4, Numbers of visiting times for selected 46 subjects. Basic statistics are given in the top right corner.

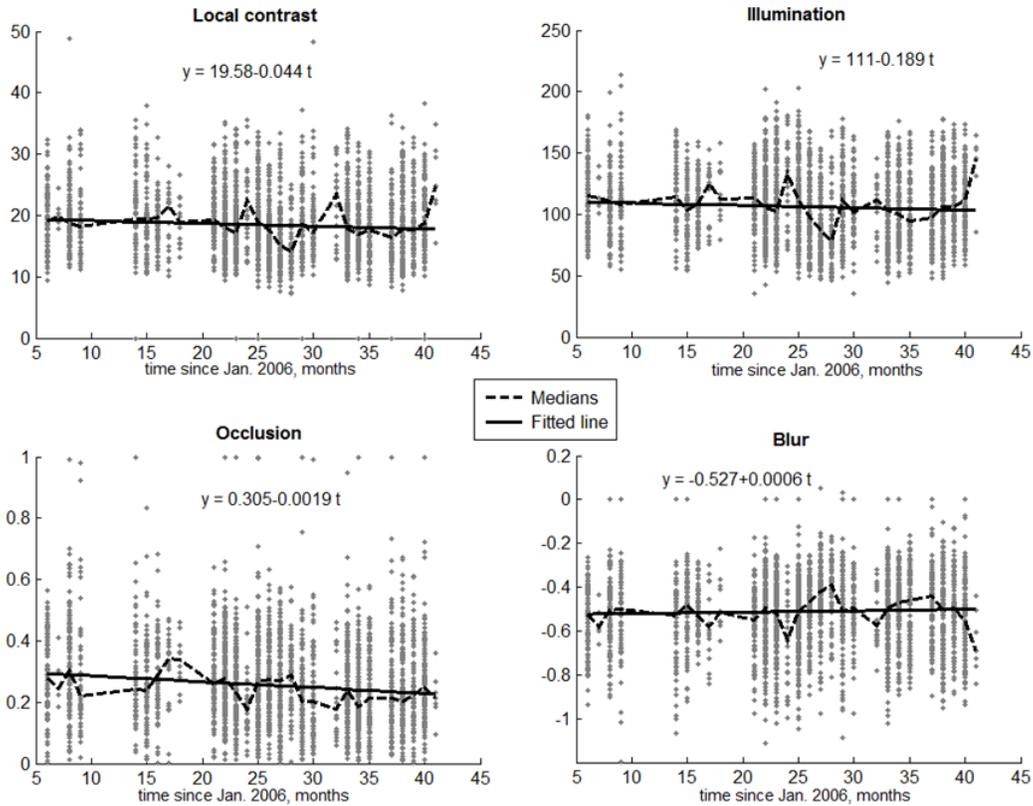


Figure 5, Four different quality factors measured for all images along the time line. Higher local contrast and illumination scores generally indicated higher quality; higher occlusion and blur scores were associated with lower quality iris images. Local contrast illumination and blurs show decreased quality over time, while occlusion shows quality improvement

### 3.3 Impact of time lapses on match scores

In order to have the convincing results, we choose 46 subjects among 244 subjects to generate the match scores. Meanwhile, we also select good quality images. Matches for images from the same day are excluded since they tend to produce good match scores due to highly correlated experimental conditions during acquisition. Thus, the analysis for all images from 46 subjects is based on 67932 match pairs and for good images was based on 27093 match pairs. Subsequent analyses show results using (a) all images from the 46 subjects and, (b) only the good quality images from the 46 subjects. Figure 6 shows the distributions of genuine match scores from the modified Masek iris recognition algorithm versus the impostor match for different time-lapsed values while Figure 7 shows the distribution of genuine match scores from commercial software VeriEye SDK. Figure 6a and 7a are scores of all images, and Figure 6b and 7b show results from the selected good images from the 46 subjects.

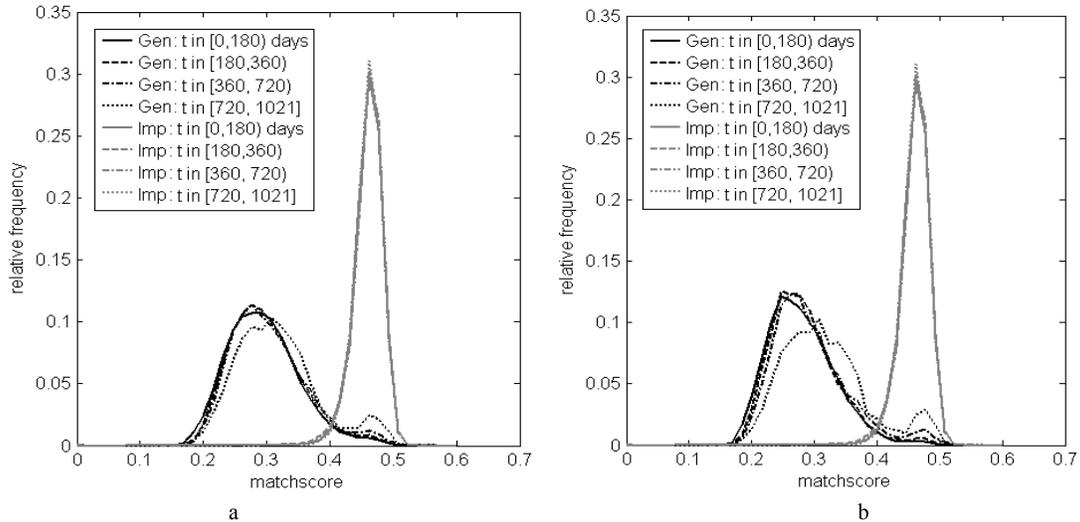


Figure 6, Histograms representing distributions of match scores from modified Masek algorithm with different time-lapsed values for 46 subjects for (a) all images and (b) only good images. For modified Masek algorithm match scores are given as Hamming distance, i.e. lower scores indicated better match

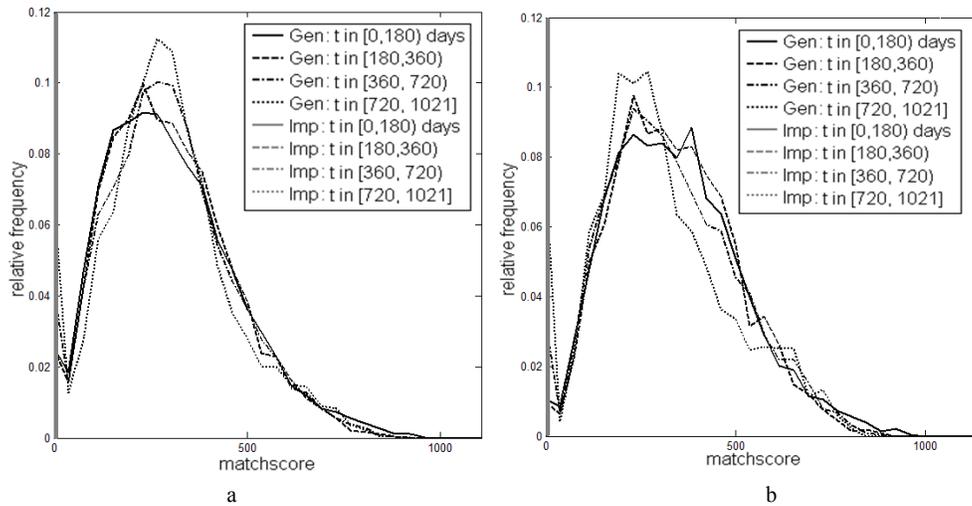


Figure 7, Histograms representing distributions of match scores from VeriEye with different time-lapsed values for 46 subjects. (a) all images, (b) Good images only. All impostor scores were zeros which are shown by the grey bar at 0 value. For VeriEye match scores similarity between two images is given, thus, higher scores corresponded to better match.

Match scores for the modified Masek algorithm represent the distance, i.e. the Humming distance, between two images

being compared. Hence, greater match scores correspond to lower similarity between iris patterns. While impostor scores are in general larger. Additionally, match scores from VeriEye represent similarity between matching images, so match scores have higher value while impostor scores have the lower ones. For this particular data set the analysis is based on approximately 100,000 randomly selected impostor match scores for the Masek encoding and around 18,000 randomly selected impostor scores for VeriEye. All of the impostor scores randomly selected for VeriEye were zeros. The iris recognition performance is assessed for different time-lapse values between matched images using detection-error-tradeoff (DET) curves and histograms of genuine and impostor match scores, where the quality of the underlying iris images is accounted for in two ways. The first is to set a threshold on the individual quality metrics and only include relatively good quality images for analysis. Thresholds were chosen to select good quality images:  $10 \leq LC \leq 25$ ,  $75 \leq IL \leq 150$ ,  $OC \leq 0.5$ ,  $SH \leq 0.3$ . These thresholds are chosen to represent the majority of the values for each particular metric and hence can be considered to be subjective. However, this seems to be adequate since the goal is to select a subset of sufficiently good images to show differences in the analyses.

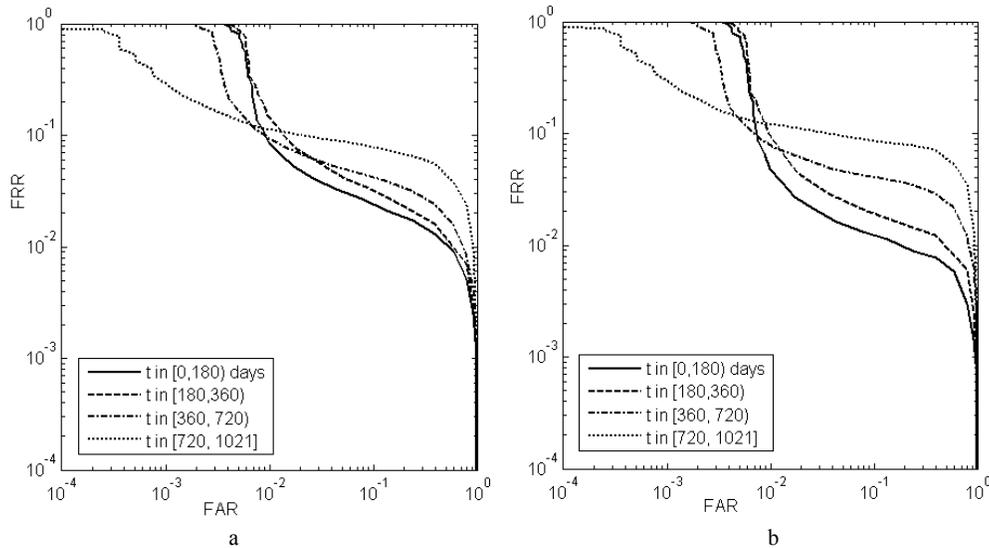


Figure 8. DET curves for modified Masek algorithm for different time lapses for 46 subjects including: (a) all images (b) good images

Table 1. Verification rates at fixed FAR %

Time lapse, days	VeriEye (at 0% FAR) <sup>a</sup>		Modified Masek (at 10% FAR)	
	All	Good	All	Good
$0 < t \leq 180$	97.5	99.1	97.3	98.7
$180 < t \leq 360$	97.5	99.0	96.6	97.9
$360 < t \leq 720$	95.9	96.8	95.4	95.7
$720 < t \leq 1021$	93.3	92.7	91.5	91.6

Compared to the histogram plots for all images in Figure 6a and 7a, Figure 6b and 7b with good quality images show a much clearer separation of those match scores corresponding to image pairs with at least a 360-day time-lapse, especially with at least 720-day time-lapse, compared to the short time lapse histograms for both algorithms. In the histograms for good images there are also an increasing (over time) portion of match scores in the range of typical impostor scores values (right peaks in Figure 6, left peaks (zero values) in Figure 7). Genuine matches show consistently less similarity. DET curves for Masek encoding shown in Figure 8 showed that for a fixed FAR value, FRR consistently increases with the increase of the time-lapse.

Table 1 shows verification rates at FAR fixed at 0% for VeriEye and 10% level for Masek for two cases (all images and only good images) for four time lapse groups. All verification rates decrease with the increase of the time lapse. In addition, the verification rates for the good images show similar effect, which suggests that while quality varies over time, it may be masking the effect rather than contributing to the effect.

### 3.4 Regression of match scores

The absence of collinearity among the predictors is essential in estimating the individual effects (and their significance) of predictors which otherwise may be hidden. Hence, we study the effect of iris time lapse on the match scores using regression models. All of the quality metrics show very low pair-wise correlation and, therefore, are considered independently. Results of the robust regression fit are shown in Fig. 9 for modified Masek method match scores and 10 for VeriEye match scores (Figure 9a and 9b are for all images and good images, respectively). Regression models are built with and without quality factors. In both cases, time lapses show the high significance in explanation of the match scores (p-value < 0.0001). All quality factors for both cases are significant (p-values for each quality metric for at least one of the images in the pair were less than 0.00001) for both modified Masek and VeriEye algorithms. The significance of quality factors in regression for good images indicates the presence of some variability in match scores due to quality factors.

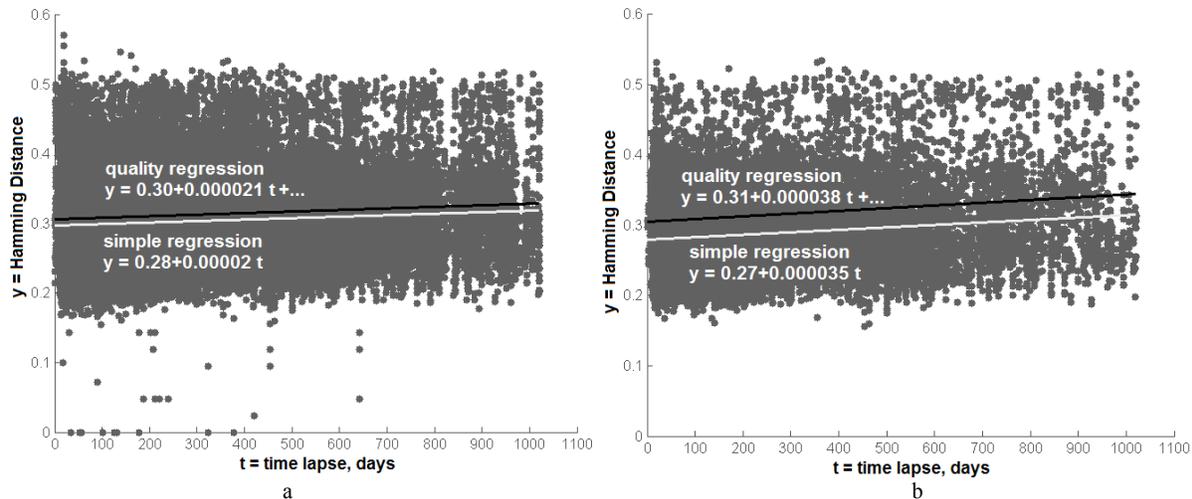


Figure 9, Robust regression fit for modified Masek match scores versus time lapses from 46 subjects with and without quality factors for: (a) all images, (b) good images.

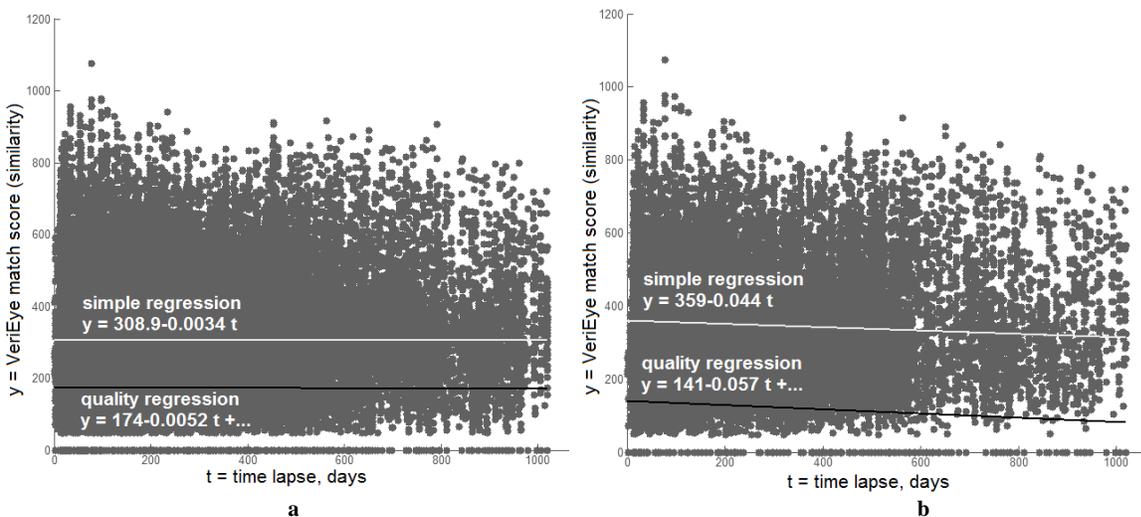


Figure 10, Robust regression fit for match scores from VeriEye versus time lapses from 46 subjects with and without quality factors for: (a) all images, (b) good images.

For modified Masek algorithm both plots in Fig. 9(a) and 9(b) show upward sloped regression lines for simple and quality regression, indicating that decreasing similarity between matched images with the increase of the time lapse

between them. Regression analysis from VeriEye matching produce downward sloping curves for both choices of images (all and good) and for both regression types (simple and quality-adjusted) indicating the decreased similarity between matched images over time. For both modified Masek and VeriEye results, the plots are constructed using only good quality images resulted in the regression lines with higher time effect (magnitude of slopes) compared to those using all images. For VeriEye this increase in the time effect is 10-fold.

Table 2 summarizes these changes in the match scores as percentage of changes per year with respect to the average score for lapses within one month. The changes are computed for both algorithms, all images and good images and two types of regressions (simple and quality adjusted).

The most significant and definite change in the match scores for both algorithms is observed for images with at least 720 days (or around 2 years) time-lapse, which becomes most evident when the analysis is based on good images (selected for specified thresholds for the four quality metrics).

Table 2, Annual change in match scores, in % per year

Regression type	VeriEye		Modified Masek	
	All	Good	All	Good
Simple	2.56%	4.47%	-0.4%	-4.47%
Quality adjusted	2.43%	4.12%	-0.6%	-5.80%

#### 4. CONCLUSION

Our study presents evaluations on effects of elapsed time on iris recognition performance taking into consideration the four factors of iris image quality by utilizing a larger number of iris images taken over two years from a legacy database at Clarkson University. Robust regression models are built with and without quality factors to evaluate the degradation of recognition performance for elapsed time. The modified Masek algorithm and commercial VeriEye SDK are used for generating match scores for iris images. In both cases, time lapses showed high significance in explanation of the match scores ( $p$ -value  $< 0.0001$ ). Performance for VeriEye decreased from 97.5% TAR at 0% FAR for time lapse less than 180 days compared to 93.3% TAR at 0% FAR for time intervals greater than 720 days. Similar decrements can be seen from the performance implemented by the modified Masek algorithm. Experimental results help confirm the regression on iris recognition performance with increased elapsed time. These results also reveal the significance of quality factors in iris recognition regression indicating the variability in match scores. Hence, our experiments can be useful for offering prediction scheme for iris recognition performance based on the learning of impact on time lapse for iris images and also of the applicability of the iris quality metric as auxiliary information to supplement iris recognition systems.

While we are controlling for quality in the analysis, it is possible to have extra factors that the quality metrics we have utilized are not adequately accounting for poor-quality images. Additional factors, e.g. different pupil dilations, the degradation from equipment, changes in data acquisition procedure, could be contributed to the impact of system performance as well. The complex inter-actions between different factors may confound the true presence of iris template aging. Moreover, it is also noticed that the regression rate is depended on matchers or data, i.e. different matchers or data could have different regression performances. This indicates that learning of degradation is needed for different systems and datasets. Our work in this paper helps provide a concept for analyzing the degradation from the database with certain matchers. Hence, our future work will explore more quality factors and better performance examiner to evaluate the impact of time lapse on iris performance.

#### REFERENCES

- [1] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, 21-30 (2004)
- [2] K. Miyazawa, K. Ito, T. Aoki, K. Kobayashi, and H. Nakajima, "An Effective Approach for Iris Recognition Using Phase-Based Image Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, 1741-1756 (2008)
- [3] P. Tome-Gonzalez, F. Alonso-Fernandez, and J. Ortega-Garcia, "On the Effects of Time Variability in Iris Recognition," in *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, 2008. BTAS 2008*, 1-6 (2008)

- [4] S. Baker, K. Bowyer, and P. Flynn, "Empirical Evidence for Correct Iris Match Score Degradation with Increased Time-Lapse between Gallery and Probe Matches," in *Advances in Biometrics*, vol. 5558, Springer Berlin / Heidelberg, 1170-1179(2009)
- [5] S. P. Fenker and K. W. Bowyer, "Experimental evidence of a template aging effect in iris biometrics," in *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, 232-239 (2011)
- [6] P. Grother and E. Tabassi, "Performance of Biometric Quality Measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, 531-543 (2007)
- [7] E. Tabassi, P. Grother, and W. Salamon, "IREX II - IQCE Iris Quality Calibration and Evaluation: Performance of Iris Image Quality Assessment Algorithms," *NIST Interagency Report 7820*, (2011)
- [8] N. Sazonova, S. Schuckers, P. Johnson, P. Lopez-Meyer, E. Sazonov, and L. Hornak, "Impact of out-of-focus blur on iris recognition," in *SPIE Defense, Security, and Sensing*, 80291S-80291S-8 (2011)
- [9] F. Hua, P. Johnson, N. Sazonova, P. Lopez-Meyer, S. Schuckers, and A. Ross, "Impact of Out-of-focus Blur on Face Recognition Performance Based on Modular Transfer Function," to be presented in *The 5th IAPR International Conference on Biometrics*,(2012).
- [10] E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A*, vol. 7, no. 10, 2032-2040 (1990)
- [11] J. Wan, X. He, and P. Shi, "An Iris Image Quality Assessment Method Based on Laplacian of Gaussian Operation," in *IAPR Conference on Machine Vision Applications*, (2007)
- [12] L. Masek, "Recognition of Human Iris Patterns for Biometric Identification," (2003)
- [13] "Multimodal Biometric Dataset Collection, Clarkson University," *Center for Identification Technology Research (CITeR)*. [http://www.citer.wvu.edu/multimodal\\_biometric\\_dataset\\_collection\\_clarksonuniv\\_release2](http://www.citer.wvu.edu/multimodal_biometric_dataset_collection_clarksonuniv_release2).