Comparison of Quality-Based Fusion of Face and Iris Biometrics

P. A. Johnson, F. Hua, and S. Schuckers

Abstract—**Multimodal systems have been used for the increased robustness of biometric recognition tasks. A unique strength of multimodal systems can be found when presented with biometric samples of degraded quality in a subset of the modalities. This study looks at the effect of quality degradation on system performance using the Q-FIRE database. The Q-FIRE database is a multimodal database composed of face and iris biometrics captured at defined quality levels, controlled at acquisition. This database allows for assessment of biometric system performance pertaining to image quality factors. Methods for measuring image quality based on illumination conditions are explored as well as strategies for incorporating these quality metrics into a multimodal fusion algorithm. This paper provides further evidence in a unique dataset that utilizing sample quality metrics into the fusion scheme of a multimodal system improves system performance in non-ideal acquisition environments.**

I. Introduction

MULTI-BIOMETRIC fusion, a method of incorporating information from multiple biometric sources, has become a common area of research in the realm of biometric recognition technology. Fusing information in biometric recognition tasks is a classifier combination problem that can take on a number of forms; given the availability of multiple expert opinions from a single feature representation, multiple feature representations from a single sample, and/or multiple independent samples available from individual classes. It has been shown that the fusion of decisions obtained from multiple experts will often outperform the decision of a single best expert [1].

There are a number of techniques available to perform information fusion in biometric systems. Of these techniques, two main categories are defined: fixed rules and classification techniques. Fixed rules are those that do not require any training and consist of sum, product, min, max, majority voting, etc. The sum rule has commonly been accepted as the highest performer in this category [1]. Further classification within the fixed rule category reveals the distinction between benign and severe rules. Benign rules, (e.g. sum and min) will allow for greater variability within the data, which comes about through the presence of noise. The more severe rules, (e.g. product and max) will often produce an undesirable veto effect in the presence of noise [2], however, if noise is low and there is low correlation between sources, the product rule has been found to outperform the sum rule [2], [3].

When component classifiers are incorporated in fixed rule fusion schemes, the effect of inter component accuracy variation comes into question. Wang et al. [4] studied this effect by incorporating iris and face biometric traits in both non-weighted and weighted sum rule fusion, showing that applying weights to each source before combination can improve performance. This technique is useful, in that it gives more decision power to the higher performing component classifiers. It has been shown in general that classification-based fusion techniques outperform fixed rules, in that a priori probability estimates allow for more informed decision making [2], [5]. The optimal classification rule would then be the Bayesian decision rule, which assigns a pattern to the class with the largest posterior probability, assuming accurate estimates of the a priori probabilities are obtained [2]. For score level fusion, the generative Bayesian approach typically takes the form of the likelihood ratio decision [5-8], while discriminative approaches, such as logistic regression [9], support vector machines (SVM) [10], and the multi-layer perceptron (MLP) [4] are also commonly used.

In multi-biometric recognition systems, the use of auxiliary information in the fusion strategy can play an important role in optimizing performance. Quality measures acquired from samples of each biometric modality can prove useful as auxiliary information to boost performance. Sample quality measures have been used in many forms in biometric systems. Aguilar et al. [11] fused match scores obtained from multiple matching algorithms for fingerprints. Quality scores extracted from the images were used to adjust the weighting between matchers to favor the more robust matcher at lower image quality. Many other studies have explored quality controlled multimodal fusion, combining traits such as fingerprint and signature [10], face and speech [12], as well as iris and fingerprint [7].

As discussed above, product rule fusion can be optimized when the correlation between sources and the noise in the data are both low. The first condition can be achieved through the use of multiple independent biometric modalities. The second condition can be optimized through the estimation and exploitation of the class conditional a priori probabilities. With the likelihood ratio statistic as the similarity score, corresponding match scores and quality measures extracted from training samples can be used to estimate the class conditional joint probability densities [7], [13]. Since these densities correlate match scores and quality measures, noise due to quality factors can be accounted for and thus the affect

This material is based upon work supported by the National Science Foundation under Grant Nos. 0325333 & 0325640, the Department of Homeland Security, and the Center for Identification Technology Research.
The authors are with Clarkson University, Potsdam NY, U.S.A (e-mail: johnsopa@clarkson.edu; huaf@clarkson.edu; sschucke@clarkson.edu).

978-1-4577-1359-0/11/$26.00 ©2011 IEEE

of the noise on performance can be diminished. The result is that a likelihood ratio statistic can be determined for each modality through the use of match score and quality measure joint densities which can be fused with the product to give a boost in classification performance over more primitive methods.

The focus of this paper is the analysis of fusion strategies with varying levels of complexity, utilizing the publicly available bi-modal face and iris Q-FIRE database [14]. This database is well-suited for evaluation of such schemes given the multitude of quality factors present, i.e. change in resolution, illumination conditions, focus blur, motion blur, off-angle gaze, and occlusion. For this paper, we will specifically explore fusion schemes as illumination is varied for both face and iris. A quality measure for each modality, relating to the illumination condition is computed, and empirical results presented, demonstrating the benefit to incorporation of quality into a fusion strategy.

The affect of illumination conditions on biometric recognition system performance has been explored in literature, particularity with regard to face recognition. Georghiades et al. [15] proposed a method to accommodate for variable lighting conditions for face recognition using illumination cone models. The illumination source can also have an effect on face skin color, impacting recognition performance [16]. Iris recognition has also been shown to be impacted by illumination conditions, where performance often decreases for low contrast images [17].

The remainder of the paper is organized as follows. Section II presents the proposed quality measures. Section III outlines the methods employed in this study, including match score generation and fusion strategies implemented. Section IV describes the experiments performed. Finally, Section V gives the conclusions.

II. Quality Measures

Variation in illumination conditions poses a significant problem for the face recognition task [18]. Factors such as illumination direction and intensity of the light source can severely alter the appearance of an individual's face and subsequently weaken genuine match scores. It is suggested in [14], that a useful measure of face image quality is the contrast of the skin area of the face. In the Q-FIRE database, a good contrast was sought through the use of a high lighting condition. Low light resulted in a low contrast value, reducing the amount of revealed detail. Last, a light source at 90°, resulting in a shadow across half the face, gives a high contrast, and is the most challenging for recognition [14][14]. The high lighting condition is achieved using overhead florescent lights in addition to two 500 Watt studio lamps, diffused with umbrellas, placed 30° off center. The low light condition is achieved with only the overhead florescent lights. The shadow condition is produced with all lights turned off, with the exception of one 500 Watt studio lamp placed at 90° to the side of the subject.

A complementary quality measure for the iris is defined as the average contrast values between sclera-iris and iris-pupil boundaries of the eye. The two contrast values are averaged to give a single measure. The three different levels of illumination are set by different numbers of LED-based NIR light spots, placed on a portal, 2 feet in front of the subject. The high, medium, and low conditions are achieved with 8, 6, and 4 light spots respectively. The quality levels corresponding to illumination conditions for this database are defined and presented in [14].

III. Methods

A. Matching

For the generation of face match scores, the FaceIT SDK is used. Iris match scores are then generated using a modified version of the Masek open source software [19], substituting a segmentation algorithm based on the circular integrodifferential operator proposed by Daugman [20], [21]. This segmentation method was able to be optimized to produce adequate results on the less constrained Q-FIRE database.

B. Fusion

Starting from the most simple strategy, the fixed sum rule is implemented for score level fusion. This method assumes no prior knowledge of class conditional densities. Given the K component score vector $\boldsymbol{x} = [x_1, x_2, \dots x_K]$, corresponding to the K matchers, the fused sum of scores $S_f(\boldsymbol{x})$ is determined as

$$S_f(\boldsymbol{x}) = \frac{1}{K}\sum_{k=1}^{K} x_k. \qquad (1)$$

It is important that the scores of both modalities be normalized to a specific range to avoid, as much as possible, a bias toward one of the modalities. In this case, all scores are pre-normalized to the range [0,1] using min-max normalization.

The second fusion method implemented is the likelihood ratio statistic as described by Nandakumar et al. [13]. The likelihood ratio statistic requires a training step, which consists of the density estimation of the genuine and impostor distributions, denoted by $f_{gen}(\boldsymbol{x})$ and $f_{imp}(\boldsymbol{x})$. These densities are modeled with Gaussian mixture models (GMM), using an independent training set of match scores, where no subjects included in the training set were used to construct the match score set for evaluation. The likelihood ratio statistic is then defined as the ratio of estimated densities at a given score vector $\boldsymbol{x}$, as shown in (2).

$$LR(\boldsymbol{x}) = \frac{\hat{f}_{gen}(\boldsymbol{x})}{\hat{f}_{imp}(\boldsymbol{x})} \qquad (2)$$

The Neyman-Person theorem states that when testing a hypothesis H_0 against H_1, $P(LR(\boldsymbol{x}) \geq \eta | H_0) = \alpha$ is the most powerful test at false accept rate α, for some threshold η, where H_0 corresponds to an impostor and H_1 corresponds to a genuine user [13].

For the third fusion method, the quality measures are exploited in a quality-based likelihood ratio method. This method modifies (2) to incorporate quality measure $\boldsymbol{q}$ and give the following quality-based likelihood ratio statistic:

$$QLR(\boldsymbol{x},\boldsymbol{q}) = \frac{\hat{f}_{gen}(\boldsymbol{x},\boldsymbol{q})}{\hat{f}_{imp}(\boldsymbol{x},\boldsymbol{q})}. \quad (3)$$

In this case, instead of estimating conditional joint match score densities, individual conditional densities joining match scores and quality measures are estimated for each of the K modalities and combined via the product as shown in (4).

$$\hat{f}(\boldsymbol{x},\boldsymbol{q}) = \prod_{k=1}^{K} \hat{f}_k(x_k, q_k) \quad (4)$$

The product is applicable here due to the assumed independence of the face and iris modalities.

IV. Experiments

The experiments conducted here aim to test each of the fusion strategies selected for evaluation. The match scores for face and iris were each generated and then the fusion strategies were implemented through the methods described in Section III. The performances of all strategies are then presented for comparison.

A. Database

The database used for this study consists of a subset of the Q-FIRE database. Since the focus of the study was evaluation of fusion strategies based on quality factors related to illumination conditions, images were selected at a single distance of 5 feet using the frame with the best focus for each of the three lighting conditions. This was to avoid the influence of any other quality factors on the current evaluation. The database was subsequently divided into a training set and an evaluation set. A total of 173 subjects are included in the full set, which is divided approximately in half, including 87 subjects in the training set and 86 in the evaluation set. The number of images, as well as the number of genuine and impostor match scores obtained from those images are summarized in Table I.

TABLE I
Summary of Q-FIRE Data Subset Used for Evaluation of Fusion Algorithms

Subset	Number of Subjects	Number of Images / modality	Number of Genuine Match Scores / modality	Number of Impostor Match Scores / modality
Training	87	261	1044	89784
Evaluation	86	258	1032	87720

B. Results

The performance metrics implemented in this study are false accept rate (FAR), which is the percentage of impostors that are wrongfully accepted at a given acceptance threshold and genuine accept rate (GAR), which is the percentage of genuine users that are rightfully accepted at a given acceptance threshold. Plotting FAR on the x-axis and GAR on the y-axis as the acceptance threshold is varied, forms the receiver operating characteristic (ROC) curve.

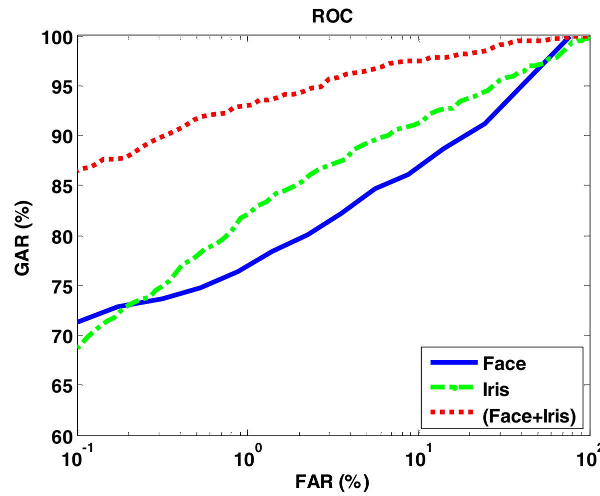

Fig. 1. Performance comparison between individual modalities and sum rule fusion.

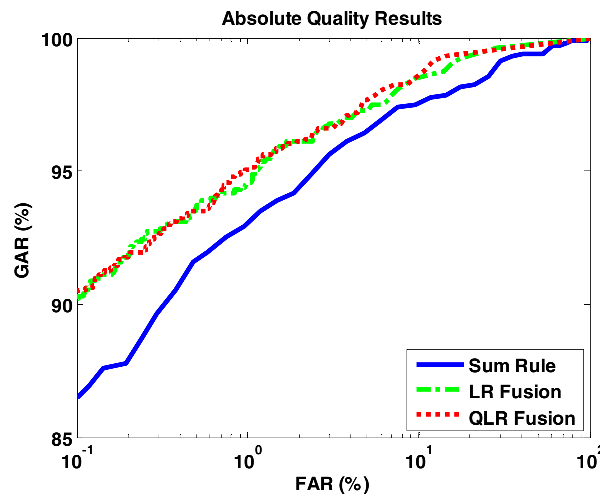

Fig. 2. Comparison of multimodal fusion strategies, incorporating absolute quality measures for the quality-based likelihood ratio fusion, obtained from the probe image.

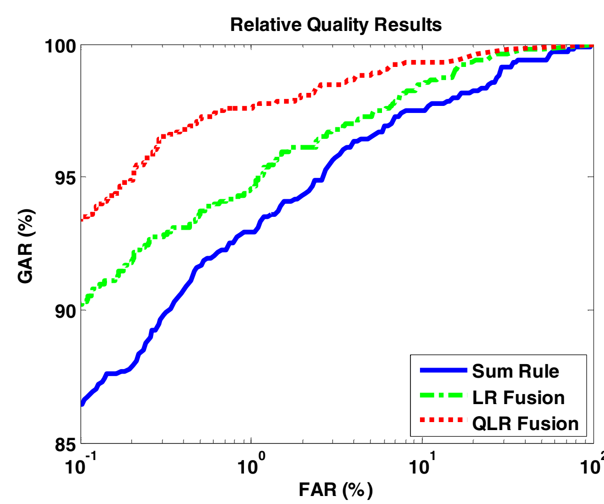

Fig. 3. Comparison of multimodal fusion strategies, incorporating a relative quality measure for the quality-based likelihood ratio fusion, obtained from both gallery and probe images.

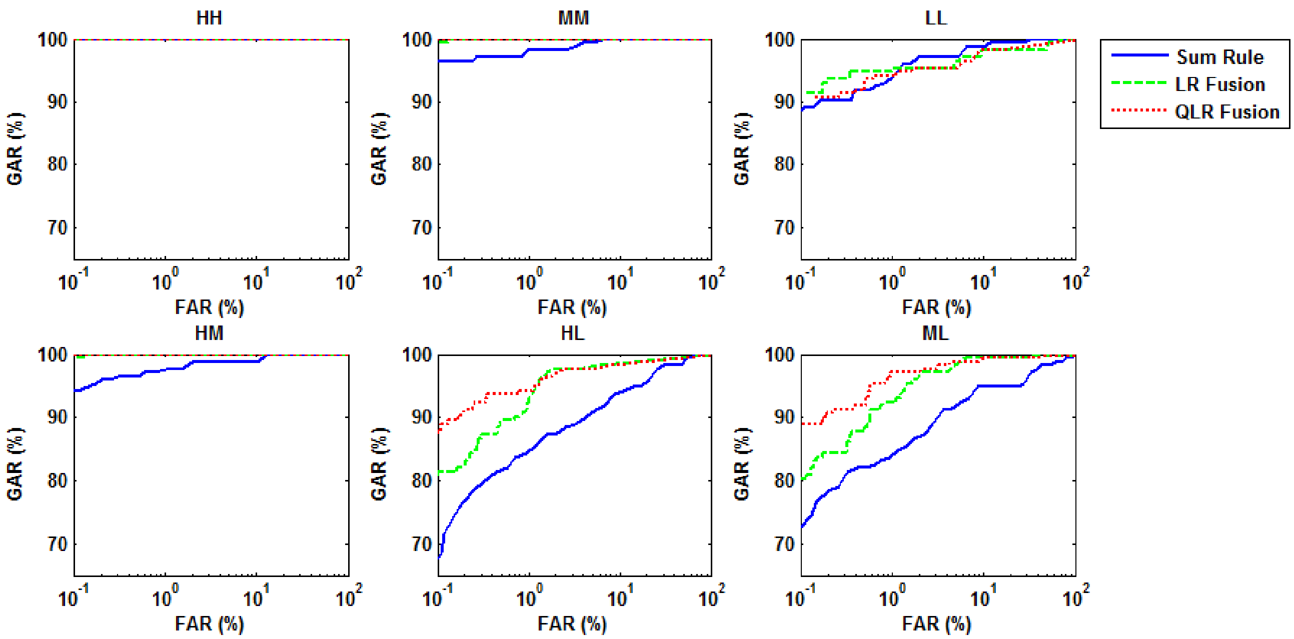

Fig. 4. Comparisons of fusion strategies by looking at performance, using relative quality, at specific pairs of enrollment/verification conditions. Subplots are labeled by condition pairs, where H, M, and L correspond to high, medium, and low quality respectively.

It is first shown in the ROC curves of Fig. 1, that the basic sum rule fusion of the face and iris match scores outperforms either one of the individual modalities alone. The overall results for the three fusion strategies are presented in Fig. 2. In this scenario, the quality-based likelihood ratio is driven by an absolute quality measure, which is simply the quality measure taken from the probe image, i.e. the sample captured at authentication. It is shown from the lack of performance boost, as seen in Fig. 2, that this absolute quality measure is of little benefit to the overall performance of the system. This is most likely due to the uncertainty of the quality of the gallery image, i.e. the sample captured and stored at enrollment. Therefore, a more useful indication of quality would be a relative measure, comparing the difference in quality between the two images being matched.

TABLE II
COMPILATION OF ERROR RATES FOR ABSOLUTE AND RELATIVE QUALITY MEASURES

Fusion	Absolute Quality		Relative Quality	
	EER	GAR (FAR=0.1%)	EER	GAR (FAR=0.1%)
Sum Rule	3.8%	86.5%	3.8%	86.6%
LR	3.2%	90.2%	3.2%	90.3%
QLR	3.3%	90.3%	1.9%	93.3%

For the analysis of a relative quality measure, quality measures are taken from both gallery and probe images and the absolute difference between the measures is used for incorporation into the quality-based fusion strategy. The results of this method, shown in Fig. 3, indicate a significant boost in performance. It is therefore evident that a relative quality measure, estimated from both images being matched, allows for a greater performance increase using quality in multimodal fusion. For this database, the quality-based likelihood ratio method was the top performer with an equal error rate (EER) of 1.9%. Associated error rates for the overall results are presented in Table II, including the GAR value at an FAR of 0.1%.

TABLE III
COMPILATION OF ERROR RATES BY QUALITY CONDITION FOR RELATIVE QUALITY MEASURES. HIGH, MEDIUM, AND LOW CONDITIONS ARE INDICATED BY H, M, AND L RESPECTIVELY. GAR EVALUATED AT A FAR OF 0.1%

Condition	Sum Rule		LR		QLR	
	EER	GAR	EER	GAR	EER	GAR
HH	0%	100%	0%	100%	0%	100%
MM	1.8%	96.5%	0.07%	99.5%	0.01%	100%
LL	3.0%	88.6%	4.6%	91.3%	4.8%	89.0%
HM	1.9%	94.0%	0.07%	99.4%	0.02%	100%
HL	7.5%	67.6%	2.4%	80.2%	2.4%	87.2%
ML	6.9%	72.7%	2.9%	80.2%	2.3%	88.4%

In Fig. 4, the results of are broken up by quality factors. This shows how the performance decreases as sample quality is degraded. The scenario where both gallery and probe images are of high quality achieves 0% error for all fusion methods, while the scenario where the gallery images are high quality and the probe images are low quality gets the worst result of 2.4%, 2.4%, and 7.5% EER for QLR, LR, and sum rule fusion respectively. In this case, the relative quality measures are implemented. It should be noted that the greatest improvements in performance, shown in Fig. 4, occur when one of the two acquisition environments is good. This is seen in the high/low and medium/low scenarios. Conversely, the low/low scenario does not show much improvement, as a relative quality measure cannot account for both samples being of low quality. The specific values for EER and GAR at an FAR of 0.1% for each condition are presented in Table III.

V. CONCLUSION

Analysis of these results shows how performance of a multimodal fusion algorithm can be affected by image quality degradation. It is also shown that through the incorporation of quality measures in the fusion scheme, the negative effect of poor quality samples can be diminished. An approach to incorporating quality measures related to illumination conditions at acquisition into multimodal fusion has been presented using the Q-FIRE database. Future work for this study will be to exploit more of the quality factors represented by the Q-FIRE database, such as blur, resolution, and occlusion to further evaluate quality-based multimodal fusion algorithms.

REFERENCES

[1] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern Analysis & Applications*, vol. 1, no. 1, pp. 18-27, Mar. 1998.

[2] D. M. J. Tax, M. van Breukelen, R. P. W. Duin, and Josef Kittler, "Combining multiple classifiers by averaging or by multiplying?," *Pattern Recognition*, vol. 33, no. 9, pp. 1475-1485, Sep. 2000.

[3] O. Fatukasi, J. Kittler, and N. Poh, "Quality Controlled Multimodal Fusion of Biometric Experts," in *Progress in Pattern Recognition, Image Analysis and Applications*, vol. 4756, L. Rueda, D. Mery, and J. Kittler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 881-890.

[4] Y. Wang, T. Tan, and A. K. Jain, "Combining Face and Iris Biometrics for Identity Verification," in *Audio- and Video-Based Biometric Person Authentication*, vol. 2688, J. Kittler and M. S. Nixon, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 805-813.

[5] S. C. Dass, K. Nandakumar, and A. K. Jain, "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems," in *Audio- and Video-Based Biometric Person Authentication*, vol. 3546, T. Kanade, A. Jain, and N. K. Ratha, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1049-1058.

[6] S. Prabhakar and A. K. Jain, "Decision-level fusion in fingerprint verification," *Pattern Recognition*, vol. 35, no. 4, pp. 861-874, Apr. 2002.

[7] K. Nandakumar, Yi Chen, A. K. Jain, and S. C. Dass, "Quality-based Score Level Fusion in Multibiometric Systems," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006, vol. 4, pp. 473-476.

[8] R. N. Rodrigues, L. L. Ling, and V. Govindaraju, "Robustness of multimodal biometric fusion methods against spoof attacks," *Journal of Visual Languages & Computing*, vol. 20, no. 3, pp. 169-179, Jun. 2009.

[9] J. Kittler et al., "Quality dependent fusion of intramodal and multimodal biometric experts," in *Biometric Technology for Human Identification IV*, Orlando, FL, USA, 2007, vol. 6539, pp. 653903-14.

[10] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and Josef Bigun, "Kernel-Based Multimodal Biometric Verification Using Quality Signals." 2004.

[11] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. K. Jain, "Incorporating Image Quality in Multi-algorithm Fingerprint Verification," in *Advances in Biometrics*, vol. 3832, D. Zhang and A. K. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 213-220.

[12] Krzysztof Kryszczuk, Jonas Richiardi, Plamen Prodanov, and Andrzej Drygajlo, "Error Handling in Multimodal Biometric Systems using Reliability Measures," presented at the 13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey, 2005.

[13] K. Nandakumar, Yi Chen, S. C. Dass, and A. K. Jain, "Likelihood Ratio-Based Biometric Score Fusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 342-347, 2008.

[14] P. A. Johnson, P. Lopez-Meyer, N. Sazonova, F. Hua, and S. Schuckers, "Quality in face and iris research ensemble (Q-FIRE)," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, 2010, pp. 1-6.

[15] A. Georghiades, P. Belhumeur, and D. Kriegman, "From Few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 660, 643, 2001.

[16] J. B. Martinkauppi, M. N. Soriano, and M. H. Laaksonen, "Behavior of Skin Color Under Varying Illumination Seen By Different Cameras At Different Color Spaces," 2001.

[17] H. Proenc¸a and L. A. Alexandre, "UBIRIS□: A noisy iris image database," *Image Analysis and Processing–ICIAP 2005*, vol. 3617, pp. 970-977, 2005.

[18] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721-732, 1997.

[19] L. Masek and P. Kovesi, "Matlab source code for a biometric identification system based on iris patterns," *The School of Computer Science and Software Engineering, The University of Western Australia*, 2003.

[20] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148-1161, 1993.

[21] J. Daugman, "The importance of being random: statistical principles of iris recognition," *Pattern Recognition*, vol. 36, no. 2, pp. 279-291, Feb. 2003.