Automatic Eye Detection Error as a Predictor of Face Recognition Performance

Abhishek Dutta Raymond Veldhuis Luuk Spreeuwers

University of Twente, Netherlands.

{a.dutta,r.n.j.veldhuis,l.j.spreeuwers}@utwente.nl

Abstract

Various facial image quality parameters like pose, illumination, noise, resolution, etc. are known to be a predictor of face recognition performance. However, there still remain many other properties of facial images that are not captured by the existing quality parameters. In this paper, we propose a novel image quality parameter called the *Automatic Eye Detection Error* (AEDE) which measures the difference between manually located and automatically detected eye coordinates. Our experiment results carried out using FaceVACS recognition system and the MultiPIE dataset show that AEDE is indeed a predictor of face recognition performance.

1 Introduction

The quality of facial images is known to affect the performance of a face recognition system. A large and growing body of literature has investigated the impact of various image quality parameters on the performance of existing face recognition systems [1]. The most commonly used image quality parameters are: facial pose, illumination direction, noise, blur, facial expression, image resolution. However, some aspects of the recognition performance that cannot be explained by the existing image quality measures remain. This shows that still more quality parameters are needed to fully explain the variation in recognition performance.

In this paper, we propose a novel image quality parameter called the *Automatic Eye Detection Error* (AEDE). Automatic eye detectors are trained to return the location of two eye coordinates in a facial image. To assess the accuracy of automatic eye detectors, we use the manually annotated eye coordinates as the ground truth eye locations. The proposed AEDE measures the error in automatically detected eye coordinates. The main insight underpinning this novel image quality parameter is as follows: Automatic eye detection becomes more difficult for poor quality facial images and hence the eye detection error should be an indicator of image quality and face recognition performance. In other words, we use the knowledge of the accuracy of one classifier (i. e. automatic eye detector) as the predictor of the accuracy of another classifier (i. e. the face recognition system) when both operate on the same pair of facial images. The proposed AEDE quality measure can be seen as providing a summary of many, but not all, properties of a facial image.

This paper is organized as follows: In Section 2, we review some previous work in this area. We explain the proposed AEDE quality measure in Section 3. We describe experiments to study the relationship between AEDE and face recognition performance in Section 4.

2 Related Work

The face recognition research community has been investigating the impact of automatic eye detection error on facial image registration which in turn influences face recognition performance [5, 12, 6, 7, 9, 14, 8]. While some researchers have focused on improving the accuracy of automatic eye detectors [13], others have explored multiple ways to make face recognition systems inherently robust to facial image registration errors [10, 11].

To the best of our knowledge, no previous work has proposed the *Automatic Eye Detection Error* (AEDE) as a predictor of face recognition performance. However, [12] make a concluding remark that points in this direction. The authors mention that "a face recognition system suffers a lot when the testing images have the lower face lighting quality, relatively smaller facial size in the image, ...". They further note that "the automatic eye-finder suffers from those kinds of images too". This paper is probably the first to observe that some facial image quality parameters (like illumination, resolution, etc.) impact the performance of both face recognition systems and automatic eye detectors.

3 Methodology

Manually annotated eye coordinates are used as the ground truth for the eye locations in a facial image. Based on this knowledge of true location of the two eyes, we can assess the accuracy of an automatic eye detector. The error in automatic eye detection gives an indication of how difficult it is to automatically detect eyes in that facial image. Some of the image quality variations that make the automatic eye detection difficult also contribute towards the uncertainty in decision about identity made by a face recognition system operating on that facial image. For example: a poorly illuminated facial image not only makes eye detection difficult but it also makes face recognition harder.

Let $p^m_{\{l,r\}}$ denote the manually located left and right eye coordinates (i. e. the ground truth). An automatic eye detector is trained to locate the position of the two eye coordinates $p^d_{\{l,r\}}$ in a facial image. The error in automatically detected eye coordinates can be quantified using the Automatic Eye Detection Error (AEDE) [4] as follows:

$$J = \frac{max\{||p^m_l - p^d_l||, ||p^m_r - p^d_r||\}}{||p^m_l - p^m_r||} \tag{1}$$

Let $J_{\{p,g\}}$ denote the AEDE in a probe and gallery image pair respectively. For this probe and gallery image pair, let s^k denote the similarity score computed by face recognition system k. We divide J into L monotonically increasing intervals (based on quantiles, standard deviation of observed $J_{\{p,g\}}$, etc.): J^l where $l \in \{1, \cdots, L\}$. We paritition the set of all similarity scores S into $L \times L$ categories of genuine G and impostor I scores defined as follows:

$$G_{(l_1,l_2)} = \{S(i) : J_p(i) \in J^{l_1} \wedge J_g(i) \in J^{l_2} \wedge S(i) \text{ denotes genuine comparision}\}, \tag{2}$$

$$I_{(l_1,l_2)} = \{S(i) : J_p(i) \in J^{l_1} \wedge J_g(i) \in J^{l_2} \wedge S(i) \text{ denotes impostor comparision}\}, \tag{3}$$

where, $l_1, l_2 \in \{1, \cdots, L\}$, $J_{\{p,g\}}(i)$ denotes the normalized eye detection error (or, AEDE) in probe and gallery image respectively corresponding to i^{th} similarity score $S(i)$. The performance of a verification experiment is depicted using a Receiver Operating Characteristics (ROC) curve. The ROC curve corresponding to a particular eye detection error interval (l_1, l_2) is jointly quantified by False Accept Rate (FAR) and False Reject Rate

(FRR) defined as follows:

$$
\begin{aligned}
FAR_{(l_1,l_2)}(t) &= \frac{n(\{I_{l_1,l_2} : I_{l_1,l_2} > t\})}{n(I_{l_1,l_2}}, \\
FRR_{(l_1,l_2)}(t) &= \frac{n(\{G_{l_1,l_2} : G_{l_1,l_2} < t\})}{n(G_{l_1,l_2}},
\end{aligned} \tag{4}
$$

where, t denotes the decision threshold similarity score and $n(A)$ denotes the cardinality of set A.

Our hypothesis is that the eye detection error J defined in (1) is correlated with face verification performance defined by (4). Therefore, we expect ROC curves corresponding to different eye detection error intervals to be distinctly different from each other. Furthermore, we also expect recognition performance to degrade monotonically with increase in eye detection error.

The proposed AEDE quality measure should be used with caution because all the factors that make eye detection difficult are not necessarily always involved in making face recognition harder. For example, a facial photograph captured under studio conditions but with the subject's eyes closed is a difficult image for automatic eye detector while a face recognition system can still make accurate decisions as most important facial features are still clearly visible. Therefore, in addition to the automatic eye detection error, we need more quality parameters in order to reliably predict face recognition performance.

4 Experiments

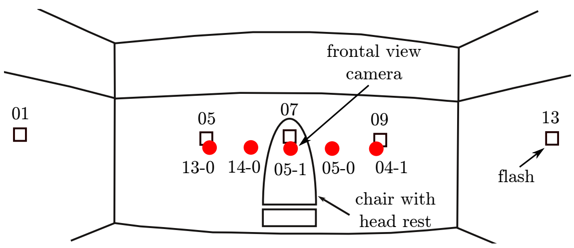

Fig. 1: MultiPIE camera and flash positions used in this paper.

In this section, we describe experiments that allow us to study the relationship between Automatic Eye Detection Error (AEDE) and the corresponding face recognition performance.

We use the facial images present in the neutral expression subset of the MultiPIE data set [3]. We include all the 337 subjects present in all the four sessions (first recording only). In our experiments, the image quality (i. e. pose and illumination) variations are only present in the probe (or, query) set. The gallery (or, enrollment) set remains fixed and contains only high quality frontal mugshots of the 337 subjects. The probe set contains images of the same 337 subjects captured by the 5 camera and under 5 flash positions (including no-flash condition) as depicted in Fig. 1. Since our gallery set remains constant, we only quantify the normalized eye detection error for facial images in the probe set J_p. Of the total 27630 unique images in the probe set, we discard 69 images for which the automatic eye detector of FaceVACS fails to locate the two eyes.

We have designed our experiment such that there is minimal impact of session variation and image alignment on the face recognition performance. We select the high quality gallery image from the same session as the session of the probe image. Furthermore, we disable the automatically detected eye coordinates based image alignment of FaceVACS by supplying manually annotated eye coordinates for both probe and gallery images. This ensures that there is consistency in facial image alignment even for non-frontal view images.

We manually annotate the eye locations $p^m_{\{l,r\}}$ in all the facial images present in our data set. Using the eye detector present in the FaceVACS SDK [2], we automatically locate position of the two eyes $p^d_{\{l,r\}}$ in all facial images. Given the manually annotated and automatically detected eye locations, we quantify the eye detection error J using (1). In Fig. 2, we show the distribution of normalized eye detection error J_p for images in the probe set categorized according to MultiPIE camera and flash identifier. The horizontal and vertical axes of Fig. 2 represent variations in camera and flash respectively. The inset images show a sample probe image with the given pose and illumination.

Now, using FaceVACS [2] recognition system, We now obtain the verification performance corresponding to each unique pair of probe and gallery images. For each verification instance, we have (J_p, s^k_{pg}) where J_p denotes the normalized eye detection error in the probe image and s^k_{pg} is the similarity score (i. e. verification score) computed by k^{th} face recognition system. Since we use only one face recognition system in our experiments, we drop the superscript k. Recall that our gallery set remains fixed to high quality images and therefore, we only consider the eye detection error of probe images. This not only simplifies the analysis and presentation of results but also simulates the conditions of a real world verification experiment. We partition the set of all similarity scores $S = \{s_{pg}\}$ into four categories based on the corresponding normalized eye detection error of the probe image J_p. If q_1, q_2, q_3 denote the $25\%, 50\%, 75\%$ quantiles of J_p, then the four categories correspond to the following interval: $J_1 = [0, q_1), J_2 = [q_1, q_2), J_3 = [q_2, q_3), J_4 = [q_3, 1)$. In Fig. 3, we show the ROC corresponding to the four intervals of J_p as shown in Table 1. The solid lines in Fig. 3 correspond to recognition performance when facial image registration is based on manually annotated eye coordinates. Section 5 describes , it will be clear that we need this result (i. e. the dotted lines o

While discussing our experiment results in Section 5, we need to rule out one possible explanation for the observed results. Therefore, in Fig. 3, we also plot the recognition performance when facial images are registered using automatically detected eye coordinates.

Table 1: Interval of J_p

Interval	Range of J_p	# Genuine	# Impostor
J_1	$[0.0, 0.0381)$	6890	1588511
J_2	$[0.0381, 0.0495)$	6890	1589314
J_3	$[0.0495, 0.0622)$	6890	1589597
J_4	$[0.0622, 1)$	6891	1585740

5 Discussion

In this paper, we set out to find if the proposed Automatic Eye Detection Error (AEDE) is a predictor of face recognition performance. Image quality parameters are very strong indicators of face recognition performance. Therefore, we first investigate if AEDE responds to controlled pose and illumination variation in facial images.

We first visually inspect the distribution of AEDE to see if it responds to the quality variations present in our data set. In Fig. 2, we show the distribution of AEDE for images in the probe set categorized according to MultiPIE camera and flash identifier. First, for the frontal camera (05_1), let us compare the distributions corresponding to frontal flash (07) and no-flash. For frontal flash, the distribution of J_p is nearly symmetric and centered around $J_p = 0.5$. For no-flash, the distribution becomes right skewed (i. e. right heavy tail)

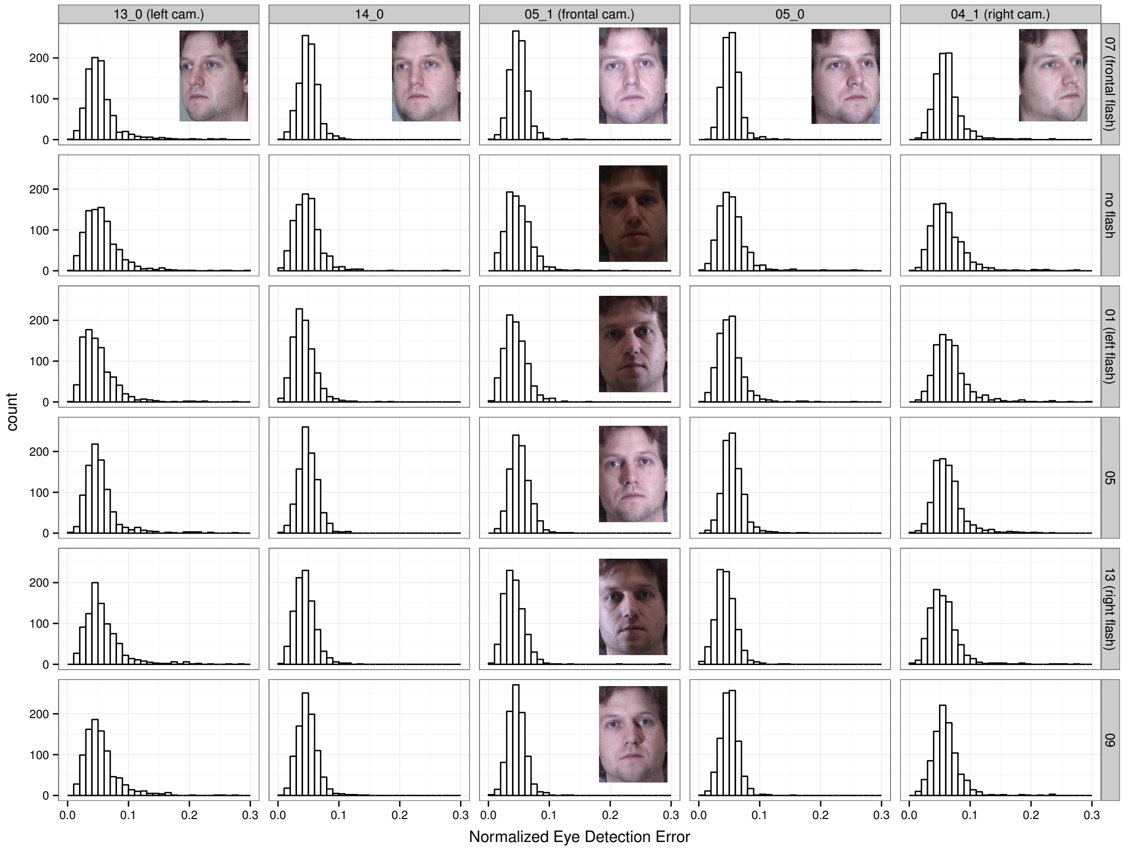

Fig. 2: Distribution of normalized eye detection error J of probe images for different pose and illumination variations from the MultiPIE data set.

indicating that many samples have high eye detection error. For other illumination variations also, we observe small increase in right skewness. This shows that the normalized eye detection error responds to illumination variations. Furthermore, higher values of AEDE corresponds to degrading illumination condition. Now let us compare the distributions for different pose variations under no-flash illumination condition. For frontal pose, the distribution of J_p is already right skewed and it becomes more heavy on the right tail as we move away from the frontal pose. This indicates that AEDE increases as the pose moves away from frontal view. Therefore, we conclude that the proposed AEDE measure responds to, at least, pose and illumination quality variations in facial images.

In Fig. 2, we show the ROC corresponding to the four intervals of the normalized eye detection error in probe image J_p. First, we discuss the four ROCs (i. e. solid lines) corresponding to facial images registered using manually annotated eye coordinates. We observe that the four intervals of J_p correspond to four distinct ROC curves. However, contrary to our expectations, the four monotonically increasing intervals of J_p do not correspond to monotonically degradaing ROC curves. For example, J_1 corresponds to the interval with lowest eye detection error but it does not correspond to the best ROC. In fact, the interval J_2 and J_3 correspond to best recognition performance. As expected, the largest eye detection error i. e. J_4 correspond to the worst recognition performance. These findings are unexpected and suggests that the normalized eye detection error has a non-linear relationship with face recognition performance. Our results further support the argument that a single metric is not sufficient to capture all image quality variations that may affect face recognition performance.

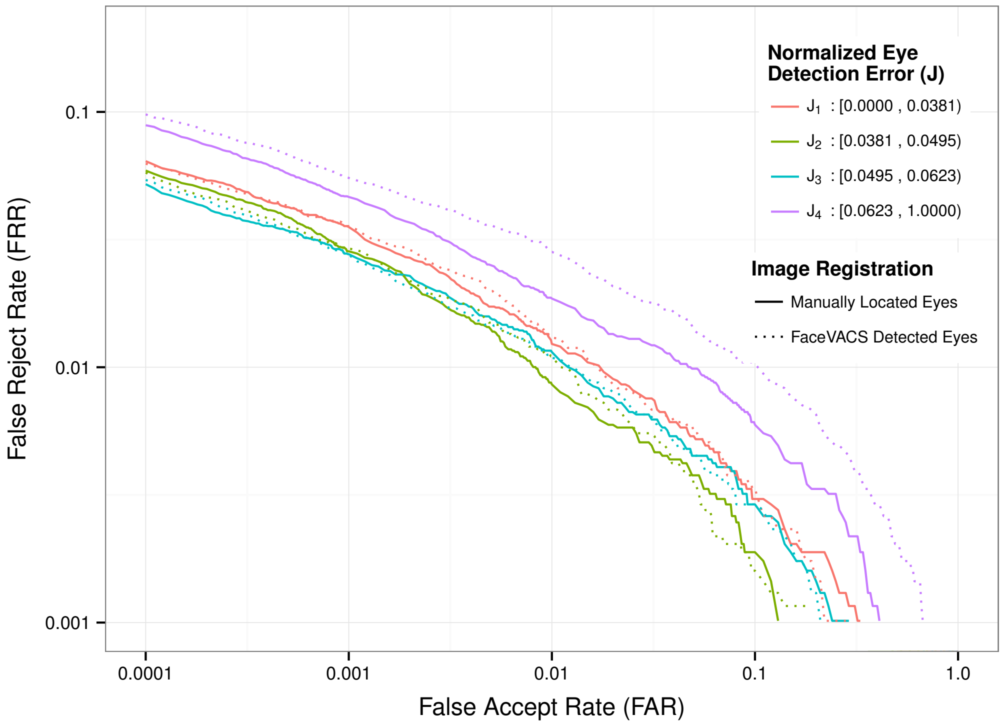

Fig. 3: Recognition performance variation for each monotonically increasing interval of normalized eye detection error J.

One could argue that the observed non-linear relationship is due to bias in the manually annotated eye coordinates and FaceVACS would behave differently if allowed to automatically register facial images. To check the validity of this argument, in Fig. 2, we plot the four ROCs (i. e. dotted lines) corresponding to facial images automatically registered by FaceVACS using its own detected eye coordinates. These ROCs also show the same trend and therefore this argument does not explain the non-linear relationship between eye detection error and recognition performance. Further work is required to determine the causes of this non-linearity.

6 Conclusion

In this paper, we have proposed Automatic Eye Detection Error (AEDE) as a predictor of face recognition performance. Our results show that AEDE has a non-linear relationship with face recognition performance and further work is required to fully understand the reasons for this non-linearity.

One of the major limitations of AEDE is that it requires manually annotated eye coordinates in order to quantify the quality of a facial image. For real time biometric applications, the manually annotated eye coordinates are usually not available. However, for forensic face recognition applications, a forensic investigator can manually annotate a small number of facial images relevant to the casework. Availability of such manual eye annotations can greatly help in quantifying the uncertainty in decision about identity using the proposed *Automatic Eye Detection Error* (AEDE) image quality measure.

The proposed eye detection error cannot capture all types of quality variations that may affect face recognition performance. For example, in a photograph containing facial image with closed eye, the eye detection error will be very high. This does not necessarily translate into a difficult verification problem. Similarly, facial expressions like smile can greatly affect face recognition performance but may not necessarily impact the performance of an automatic eye detector. Therefore, we need more quality parameters to fully quantify the variability in recognition performance.

Acknowledgements

- This work was supported by the BBfor2 project which is funded by the EC as a Marie-Curie ITN-project (FP7-PEOPLE-ITN-2008) under Grant Agreement number 238803.
- We would like to thank Cognitec Systems GmbH. for supporting our research by providing the FaceVACS software. Results obtained for FaceVACS were produced in experiments conducted by the University of Twente, and should therefore not be construed as a vendors maximum effort full capability result.

References

[1] J. R. Beveridge, G. H. Givens, P. J. Phillips, B. A. Draper, and Yui Man Lui. Focus on quality, predicting FRVT 2006 performance. In *Automatic Face Gesture Recognition, 2008. FG 08. 8th IEEE International Conference on*, pages 1–8, 2008.

[2] Cognitec Systems. FaceVACS C++ SDK Version 8.7.0, 2012.

[3] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-PIE. In *Automatic Face Gesture Recognition, 2008. FG 08. 8th IEEE International Conference on*, pages 1–8, 2008.

[4] Oliver Jesorsky, Klaus J Kirchberg, and Robert W Frischholz. Robust face detection using the hausdorff distance. In *Audio-and video-based biometric person authentication*, pages 90–95. Springer, 2001.

[5] Joe Marques, Nicholas M Orlans, and Alan T Piszcz. Effects of eye position on eigenface-based face recognition scoring. *Image*, 8, 2000.

[6] Jaesik Min, KevinW. Bowyer, and PatrickJ. Flynn. Eye perturbation approach for robust recognition of inaccurately aligned faces. In Takeo Kanade, Anil Jain, and NaliniK. Ratha, editors, *Audio- and Video-Based Biometric Person Authentication*, volume 3546 of *Lecture Notes in Computer Science*, pages 41–50. Springer Berlin Heidelberg, 2005.

[7] Terry Riopka and Terrance Boult. The eyes have it. In *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications*, WBMA '03, pages 9–16, New York, NY, USA, 2003. ACM.

[8] Yann Rodriguez, Fabien Cardinaux, Samy Bengio, and Johnny Mariéthoz. Measuring the performance of face localization systems. *Image and Vision Computing*, 24(8):882 – 893, 2006.

[9] Shiguang Shan, Yizheng Chang, Wen Gao, Bo Cao, and Peng Yang. Curse of mis-alignment in face recognition: problem and a novel mis-alignment learning solution. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 314–320. IEEE, 2004.

[10] A. Wagner, J. Wright, A. Ganesh, Zihan Zhou, H. Mobahi, and Yi Ma. Towards a practical face recognition system: Robust alignment and illumination by sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(2):372–386, Feb 2012.

[11] R. Wallace, M. McLaren, C. McCool, and S. Marcel. Inter-session variability modelling and joint factor analysis for face authentication. In *International Joint Conference on Biometrics (IJCB)*, pages 1–8, 2011.

[12] Haoshu Wang and Patrick J. Flynn. Sensitivity of face recognition performance to eye location accuracy. *Biometric Technology for Human Identification II, Proc. SPIE 5779*, pages 122–131, 2005.

[13] Peng Wang, M.B. Green, Qiang Ji, and J. Wayman. Automatic eye detection and its validation. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 164–164, June 2005.

[14] Peng Wang, Qiang Ji, and J. L. Wayman. Modeling and Predicting Face Recognition System Performance Based on Analysis of Similarity Scores. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):665–670, 2007.