Common Mistakes on Face Recognition Based on Video

V. DESPIEGEL, S. GENTRIC.
Morpho, Safran Group, (Issy-les-Moulineaux)
11 bd Gallieni, 92130 Issy Les Moulineaux, France.
vincent.despiegel@morpho.com,stephane.gentric@morpho.com

Abstract: Automatic border control e gates relies on face recognition in video streams. Two family of methods could be imagined to define a common workflow: encode all that is possible or select the best image of the stream according to a quality assessment. In gates, giving an answer as soon as possible by encoding each and every possible image seems to lead to faster gates. Is it really the case?

1 Introduction

Face Recognition has been a wide and well studied topic in machine learning and image processing for decades. The various approachs for identification and authentication, studied [ZCPR03], [JL05], have focused on comparison between still images. However, in a typical authentication scenario such as ID document checking for border control, it is natural to think of video acquisition instead of single shot image. Having video data seems to bring only advantages: more robustness to pose, to expression and even in certain conditions to illumination.

In the automatic border control scenario, it seems extremely tempting to match all the detected faces as soon as possible and to validate the authentication as soon as a score above a given threshold is available.What are the advantages and drawbacks of such a method compared to another one based on quality measures and restraining the number of comparisons to one image?

2 The Gate Scenario: How to control the False Acceptance Rate?

In Gate context, the scenario is assymetric in the way that the reference is usually the face picture contained in an ID document. On the other side, the acquisition gate leads to less control acquisition as it is usually supposed to be unsupervised. For the simplicity of the experiment, one will consider here a system with only one camera but could be adapted without any change to a multi-camera acquisition module.

2.2 Score Driven Methods vs Quality Driven Methods

Two main families of methods could be imagined to define the “behaviour” of the gate, one called score driven methods family, the other one called the quality driven methods family. With score driven methods, each input image where a face is detected may be

encoded, the features extracted and the template matched against the reference template (the one from the ID document). As soon as the matching score is above a given threshold T, the image is logged, the system open the gate's doors and the authentication is granted. If the timeout is reached without any score above the given threshold, then, the doors remain closed. This method has the advantage of simplicity, no quality is needed to select the images to be matched and the built of a gate from a face feature extractor and a comparison algorithm is straight forward.

With quality driven methods, each image where a face is detected is processed to extract some quality measurements (ICAO criteria qualities …) As soon as the fused quality is above a given threshold Q, the selected image is encoded, logged and compared to the reference template. If the matching score is above a given threshold T', the gates open the doors. Otherwise, it is a non match and no other image are tested. If the timeout is reached, the image with the best quality is encoded and the same comparison is performed. To use this kind of methodology, a quality which is a good oracle of the image quality with regards to the hit probability is necessary.

Is it really necessary to bother with quality to select the image one wants to compare with the ID document and isn't the score driven methods faster whatever happens for better or similar performances?

2.3 Mastering the False Acceptance Rate. The Log Image Dupery.

In the border control scenario, it is crucial to be able to estimate the security of the border during the lifetime of the system. As the security of the gate is determined by the False Acceptance Rate, it is necessary to be able to measure it as precisely as possible from the log data. However, it is not feasible to make the number of impostor tests needed. Thousands of impostor tests would be necessary with respect to the rules of 30 [MW02] to measure classical False Acceptance Rate of 10-3 which could be set on a classical border control. Consequently, one would like to use the log images from real passengers acquisition. It should be remarked that of course all these acquisitions are genuine ones. For the quality driven methods, as the decision on which image should be used and compared to the reference image is made without any measurement on the reference image, an impostor acquisition would lead to exactly the same selection as a genuine one (no matter whose ID document the user is bearing) So measuring the Operational False Accept Rate from log data is possible and disposing of a bunch of ID photos and log images could enable numerous impostors tests which could be performed offline and give a continuous evaluation of the operational security of the border.

One could be tempted to do exactly the same with score driven methods. However, the duration of the acquisition for a score driven method depends on whether the test is an impostor test or not. As the condition of independance between the reference image and the log image is not fullfilled, it is impossible to simulate offline impostors tests from genuine acquisition log without any bias. The question here regards the importance of this bias in term of FAR.

2.4 The Bias between the operation FAR and the offline measurement from the log.

If one estimates the FAR on the log image (from a genuine test) instead of running a real impostor test, one would compare the reference image with only one image (the log image) and determine if the score is above a given threhold T instead of comparing all the images acquired during the walk through the gate. If all the acquisitions were independent and randomly chosen, the probability that a score would be higher than the given threshold could be easily calculated. Of course, the independence condition between the acquisitions images is not at all verified but even if it is the same ID during the walkthrough, there would be variations (distance to the camera, pose, illumination, noise of the camera) which would imply variations on the scores. The number of frames used indeed has an influence on the False Accept Rate associated to a given threshold. Using an internal comparison algorithm, if one sets the threshold T to have a given operational False Accept Rate of 0.5% on the whole walkthtrough (50 images), the measured FAR on the first image would be around 0.02%.

The database is composed for the search data of 180 acquisitions until the timeout (5 fps, 10s). Concerning the reference database, it is composed of 4800 ICAO images (ID document-like acquisitions: ICAO compliant, compressed in JPG, 17ko on average). As all the acquisitions have been made before the timeout, it is possible to measure the FAR for both score driven methods and quality driven methods offline. The following results are obtained (see Figure 1).

As observed in the Figure 2, for the score driven method, there is a difference between the operational FAR and the FAR measured offline from the log image: the measured FAR is 17 times lower than the effective one and leads to a significant over evaluation of the effective security of the system.

		FAR	FRR	Mean time (genuine tests)
Score Driven Method	FAR measured on the image logged from a genuine test	0,03%	7,6%	
	Operational FAR	0,5%		5,2s
Quality Driven Method	FAR measured on the image logged from genuine tests	0,5%	4,4%	
	Operational FAR			7s

Figure 1: FRR@FAR=0.5% for quality driven and score driven methods.

For score driven methods, changing the threshold yields to different tradeoffs in terms of FAR and average acquisition duration. Concerning the quality driven methods, two level can be adjusted: the matching threshold and the quality threshold. The quality threshold is directly correlated to the average duration. A well chosen quality leads to better performances in term of FRR for quality driven methods with a shorter average duration (measured on genuine acquisitions as it is related to the number of passengers per gate per minute in an airport).

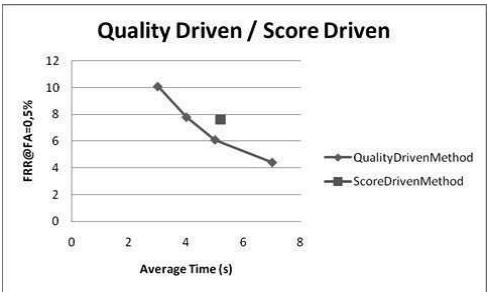

Figure 2: FRR@FAR=0.5% quality driven methods compared to score driven methods.

To conclude, even if the score driven methods intuitively seems to lead to a smaller average time, it is possible for quality driven methods to have the best of both worlds, a smaller average time and better biometrics performances. Of course the quality should be adapted to the algorithm and be a good oracle of the capability of the algorithm to accurately match such data.

3 Conclusion

Normalization issues and False Accept Rate should be taken into account too for gates scenario. The False Accept Rate in Gate Scenario is an invisible figure. In an operational point of view, the only figures that are visible to the passengers and the operators are the False Reject Rate and the time of acquisition. However, the FAR is the figure that matters for security concerns: the False Accept Rate measures the security of the border control. It is important to be able to guarantee it and estimate it regularly from operational data. Whereas score driven methods implies bias in the offline evaluation of the FAR, quality driven methods enable such offline control while giving the best of the two worlds: better biometrics performance without degrading the overall time.

References

[HDGDB12] Herold, C.; Despiegel, V.; Gentric, S.; Dubuisson, S.; Bloch, I.: Modeling, Reconstruction and Tracking for Face Recognition. Signal and Image Processing for Biometrics, édité par A. Nait-Ali and R. Fournier. Wiley-ISTE, 2012.

[JL05] Jain, A.;Li, S.: Handbook of Face Recognition, Springer-Verlag New York, Inc. Secaucus, NJ, USA 2005.

[MW02] Mansfield, A.; Wayman, J.: Best Practices in Testing and Reporting Performance of Biometric Devices, Ver 2.01, Aug 2002.

[PFSBCHMMW05] Phillips, P.; Flynn, P.; Scruggs, T.; Bowyer, K.; Chang, J.; Hoffman, K.; Margues, J.; Min, J.; Worek, W.; Overview of the face recognition grand challenge. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2005), pp. 947-954, Washington DC, 2005.

[ZCPR03] Zhao, W.; Chellappa, R.; Phillips, P.; Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys, vol. 35, no. 4, pp. 399-458, 2003.