View Based Approach to Forensic Face Recognition

Abhishek Dutta, R.T.A. van Rootseler, Raymond Veldhuis, Luuk Spreeuwers

Signals and Systems Group, Faculty of EEMCS,
University of Twente, Netherlands
`{a.dutta,r.t.a.vanrootseler,r.n.j.veldhuis,l.j.spreeuwers}@utwente.nl`

Abstract—**Face recognition is a challenging problem for surveillance view images commonly encountered in a forensic face recognition case. One approach to deal with a non-frontal test image is to synthesize the corresponding frontal view image and compare it with frontal view reference images. However, it is often difficult to synthesize a good quality frontal view image from a surveillance video because the test image is usually of low quality. In this paper, we investigate if it is useful to instead transform the reference images so that it matches the pose, illumination and camera of the surveillance view test image. This approach, also called the view based approach, ensures that a face recognition system always gets to compare images having a similar, not necessarily the frontal, view. Our results with surveillance view images captured 6 months apart (taken from the MultiPIE data set) and using five different face recognition systems show that improved recognition performance under surveillance conditions can be attained by exactly matching the pose, illumination and camera between the test and reference images.**

I. INTRODUCTION

Forensic investigators now have access to video recordings of many crime scenes – thanks to the omnipresent CCTV cameras. Such video recordings are often of very low quality and therefore rarely contribute to a strong evidence in the court of law, because forensic investigators find it difficult to compare and interpret the low quality facial images contained in these recordings. Automatic face recognition systems also have poor performance because they are fine tuned for optimal recognition performance when comparing good quality frontal view images.

One solution to deal with low quality face images is to reconstruct 3D face model from the CCTV images and synthesize the corresponding frontal view image. This strategy ensures that a face recognition system always gets to compare frontal view images, thereby ensuring optimal recognition performance. This approach is known as the model based approach. If the 3D model reconstruction is accurate and the synthesized frontal view image is of good quality, such a model based approach is known to deliver good recognition accuracy [1], [2], [3], [4]. In most forensic cases, the images extracted from the CCTV footage have surveillance view (range of azimuth $\pm 45°$, elevation $\sim 30°$) as shown in Fig.1. Therefore it is very difficult to synthesize the corresponding good quality frontal view images that can be compared to the reference image that is usually a frontal mug shot.

In the forensic context, little attention has been paid to the view based approach first examined by [6]. This approach

Fig. 1: Sample of surveillance view images commonly encountered in forensic cases (taken from MultiPIE [5])

involves adapting the test and reference images so that a face recognition system always gets to compare images under similar view – not necessarily the frontal view. The basis for the view based approach is that, given appropriate training and suitable classifiers, comparing non-frontal view facial images is no more difficult than comparing frontal view images and some face recognition algorithms (for example LBP [7]) perform equally well in both tasks. This approach has not been studied well because it is often not practical to capture reference images from all possible pose and illumination variations.

In this paper, we study the use of the view based approach for forensic cases where there is a possibility of capturing suspect reference images from desired pose and illumination using a desired camera. Our results on the MutltiPIE data set [5] shows that exactly matching pose, illumination and camera between test and reference images delivers improved recognition performance across five different types of face recognition systems.

II. RELATED WORK

A forensic evaluation case involving face recognition often involves surveillance view images. There are generally two approaches available to deal with non-frontal view (or, pose) facial images in a face comparison process using a pre-trained view based face recognition system: *a*) Model based approach *b*) View based approach .

The model based approach [1], [2], [3], [4] exploits the fact that most face recognition systems are fine tuned for optimal recognition performance when comparing frontal view images. This approach begins with reconstruction of a 3D face model from non-frontal view test image followed by synthesis of a frontal view test image (also called virtual test image) for comparison with the frontal view reference images. This

approach is applied to all the non-frontal view images present in either the test or the reference set so that a view based face recognition system only compares frontal view face images – thereby ensuring optimal recognition performance.

To the best of our knowledge, results based on the model based approach has only been reported for non-surveillance view images. In [2], the authors synthesized frontal view images corresponding to the non-frontal view images using a 3D Morphable Model (3DMM) and reported large improvement in recognition performance due to this view transformation. The results were based on good quality images captured at the eye level (i.e. non-surveillance view). More recently, [4] proposed a 3D pose normalization method based on a view based Active Appearance Model (AAM) in order to synthesize a frontal image from a non-frontal view and reported improved recognition performance on five different image data sets. Although performance improvement was reported for ± 45 pose variation, surveillance-view images were not included in the study. In [3], the authors used Structure from Motion (SFM) to infer 3D face shape information by tracking a large number of feature points in a video sequence. Again, the reported improvement in recognition performance were based on a non-surveillance view video sequence.

The authors of [6] used the view-based approach to address the problem of recognition under general viewing orientation. They partitioned the face space into view-specific regions and compared a given non-frontal test image using eigenfaces of a particular region of the face space (corresponding to the view in the test image). The basic idea was to compare face images under similar view.

In this paper, we investigate whether it is useful to apply a similar view-based approach in forensic cases where the test image is usually of very low quality. In section III, we describe the experimental setup that we used to study the performance of the model and view based approach for the surveillance view test set taken from the MultiPIE data set [5]. In section IV, we discuss the performance of five pretrained face recognition systems for this setup. Finally, based on these results, we present our recommendations for the forensic community.

III. RECOGNITION EXPERIMENT AND RESULTS

With the experiment described in this section, we want to test the performance of the model and view based approaches in a scenario commonly encountered in forensic cases. For both approaches, we evaluate the performance of the following five face recognition systems: two commercial face recognition systems denoted by A and B, Local Region PCA (LR-PCA) and LDA - I/Red (LDA-IR) [8], and Local Binary Pattern (LBP) [7] where, PCA and LDA are holistic methods while LBP is a local method. We use the value of True Positive Rate (TPR) at False Positive Rate (FPR) of 0.001 as the metric for recognition performance comparison.

Our test set (or probe) consists of surveillance view images of 249 subjects in session 01 with illumination that is frontal

with respect to the face: $(01,19\_1,18)$ [1] as shown in Fig.1. The reference set consists of frontal images of 239 subjects in session 04 with frontal illumination: $(04,05\_1,07)$. The camera and flash positions of the MultiPIE capture environment are shown in Fig.2. Note that session 01 (test set) and session 04 (ref. set) were captured six months apart [2] and therefore this experiment simulates the session variation present in real forensic cases.
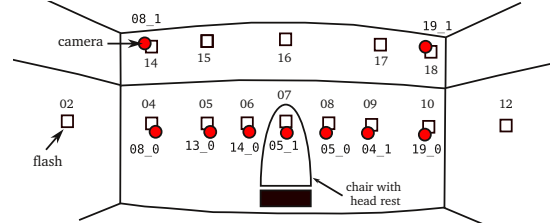

Fig. 2: Position of camera (red circles, e.g. 19_1) and flash (black squares, e.g. 18) in the MultiPIE collection room [5].

The model and view based approaches differ in the way they transform the reference images. In the following sections, we discuss the details of this process by which the reference set is transformed in these two approaches:

Model Based Approach : There are several methods to implement the model based approach [2], [3], [4]. In this paper, we use the 3D Morphable Model (3DMM) based method of [1], [2] to synthesize frontal view image corresponding to a given surveillance view image shown in Fig.1. We manually annotate 10 landmarks in the test image and then fit the Basel Face Model [9]. We then synthesize a frontal view image using the estimated pose, shape, texture and illumination. To study the effect of texture on recognition performance, we synthesize two images as shown in Fig.3. The first image contains texture from the morphable model and since [9] is based on 200 faces, it is unable to reproduce local characteristics such as moles or scars. The second image contains partial texture from the original image supplemented with morphable model texture in the occluded regions. Therefore, we observe some artifacts in the synthesized image which has also been reported in [9]. One possible reason for this artifact is the mapping of nonface (e.g. background) pixels to the model because the shape fitting process was not 100% accurate.

The result of face comparison between synthesized frontal view image and frontal view reference photograph using the five face recognition systems is shown in Fig.4a (with only morphable model texture) and Fig.4b (with partial original texture supplemented by morphable model texture). The corresponding true positive rate values (at FPR = 0.001) are shown in Table 4e.

---

[1]We use the notation (session-id,cam-id,flash-id) to denote a subset of MultiPIE data set with neutral expression

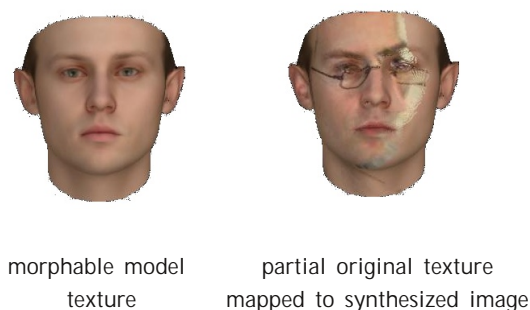[2]based on communication with an author of [5]

morphable model   partial original texture
    texture   mapped to synthesized image

Fig. 3: Synthesized frontal view image with two different types of texture

View Based Approach : Recall that in a view based approach, the reference image is chosen such that its pose closely matches the pose in the test set (i.e the surveillance view). In this paper, we investigate two scenarios relevant to real forensic cases. The first is a more ideal case where we have access to the original CCTV camera (that captured the test image) and it is possible to capture the suspect's photograph from exactly the same pose and illumination condition present in the test image. Second, is a more constrained case where we neither have access to the original CCTV camera nor are able to photograph suspects under exactly the same pose and illumination condition present in the test image. The second case is often encountered in real forensic cases.

The first case can be simulated with a reference set consisting of surveillance view images taken from session 04: `(04,19_1,18)`. This reference set not only exactly matches the pose and illumination in the test set but also matches the camera as shown in Fig.4c (top). Recognition performance for such a test and reference set is shown in Fig.4c and the corresponding true positive rate values (at FPR = 0.001) are shown in Table 4e.

To simulate the second case, we create a reference set consisting of near-surveillance view images taken from session 04 with illumination that is frontal with respect to the face: `(04,19_0,10)`. The 19_1 and 19_0 camera positions in the MultiPIE data set differ by an elevation and azimuth angle of $25.9°$ and $0.3°$ respectively as shown in Fig.2. In reality, we can more closely match the pose between test and reference images. Recognition performance for such a test and reference set is shown in Fig.4d (bottom) and the corresponding true positive rate values (at FPR = 0.001) are shown in Table 4e.

IV. DISCUSSION

For the model based approach, performance across all five systems degrades dramatically when the synthesized frontal view image contains texture from the morphable model as shown in Fig.4a. With partial texture from the original test image mapped to the synthesized frontal view image, the performance improves for commercial systems A (0.36) and B (0.13) as shown in Fig.4b which shows that, to some extent, these systems are robust to the artifacts present near the boundary of the actual and synthesized texture. On the other

hand, LR-PCA, LDA-IR (the holistic methods) and LBP (the local method) show virtually no improvement (at FPR = 0.001) in performance because they are only trained for comparing near frontal views and are also unable to handle the artifacts. These results also show that texture in the synthesized frontal view image is critical to face recognition performance in the model based approach. It is important to realize that our true positive rate values for the model based approach (for instance: TPR = 0.36 at FPR of 0.1%) are significantly lower than that reported in [1], [2] because our test set contains surveillance view images while [1], [2] used non-frontal images captured at the eye-level.

View based approach delivers improved performance across all the five face recognition systems when pose, illumination and camera match exactly between the test and reference images as shown in Fig.4c. For the test and reference set captured by different camera and having large mismatch in pose and illumination, only the commercial system A (and to some extent LBP) shows slight improvement in performance at FPR = 0.001. This reflects the capability of system A (and to some extent of LBP) to handle pose mismatch when comparing non-frontal view images.

In forensic cases, if we can synthesize good quality frontal view image with original texture, then a face recognition system robust to image synthesis artifacts (as shown in Fig.3 - right) can provide good recognition performance. However, in most real forensic cases, the test image is of very low quality and it is difficult (and often not possible) to synthesize good quality frontal view image with the original texture.

Under such a constraint, our study shows that a forensic investigator has the following two options. First, is to acquire the camera that captured the original test image (i.e. the trace) and photograph the suspects from exactly the same pose and illumination. Our results show that this approach results in improved performance across all five face recognition systems included in this study. Second, is to approximately match the pose and illumination in the test and reference images captured by different camera. Our results shows that performance of some face recognition systems (for instance system A and LBP) show a sign of improvement even if the test and reference images are captured by different camera have large mismatch in pose and illumination.

V. CONCLUSION

For a forensic evaluation case involving face recognition, our results show that the proposed view based approach delivers improved recognition performance if: *a*) it is possible to exactly match pose, illumination and camera between the test and reference set images, and *b*) you have access to a face recognition system that can compare non-frontal view images. It is still possible to attain good performance by approximately matching pose and illumination in the test and reference images captured by different camera. Our results also show that the model based approach should only be applied if: *a*) it is possible to synthesize good quality frontal view image with the original texture, and *b*) you have access to a face

recognition system that can handle artifacts caused by image synthesis techniques .

Future research could investigate how the proposed view based approach performs with non-frontal reference images synthesized by applying image synthesis techniques used in the model based approach. In real forensic cases, quite often, it is not possible to photograph the suspects and the forensic investigator has access only to frontal view (mug shot) images of the suspects. Under such a constraint, we expect the synthesized non-frontal view images to be of good quality because of the relatively better quality of frontal test images (mug shot) in the reference set. The case when we exactly match pose, illumination and camera depicts the performance achievable using photo-realistic synthesis of non-frontal view reference image.
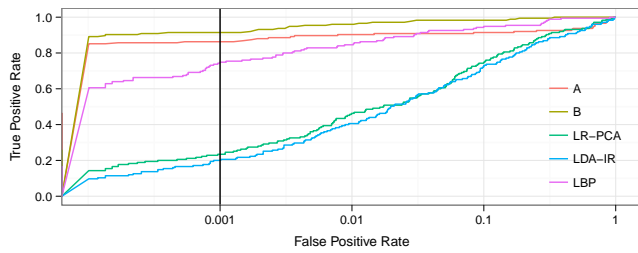
References

[1] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1063–1074, 2003.

[2] V. Blanz, P. Grother, P. J. Phillips, and T. Vetter, "Face recognition based on frontal views generated from non-frontal images," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 454–461.

[3] U. Park and A. Jain, "3D Model-Based Face Recognition in Video," in *Advances in Biometrics*, S.-W. Lee and S. Li, Eds. Springer Berlin / Heidelberg, 2007, vol. 4642, ch. Lecture Notes in Computer Science, pp. 1085–1094.

[4] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, "Fully automatic pose-invariant face recognition via 3D pose normalization," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 937–944.

[5] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Automatic Face Gesture Recognition, 2008. FG 08. 8th IEEE International Conference on*, 2008, pp. 1–8.

[6] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR 94., 1994 IEEE Computer Society Conference on*, jun 1994, pp. 84–91.

[7] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[8] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. OToole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer, "An introduction to the good, the bad, amp; the ugly face recognition challenge problem," in *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. http://www.cs.colostate.edu/facerec/algorithms/baselines2011.php, 2011, pp. 346–353.

[9] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D Face Model for Pose and Illumination Invariant Face Recognition," in *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS.*, 2009, pp. 296–301.
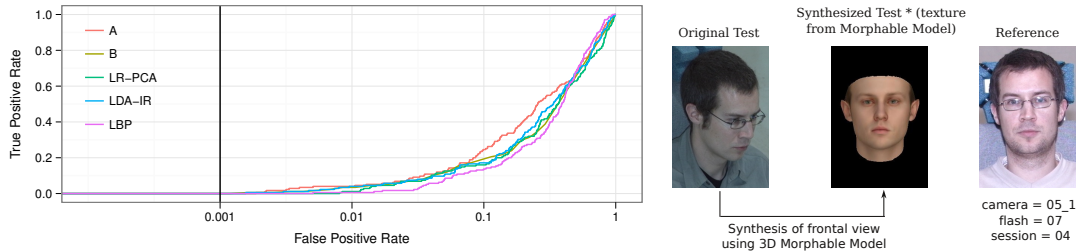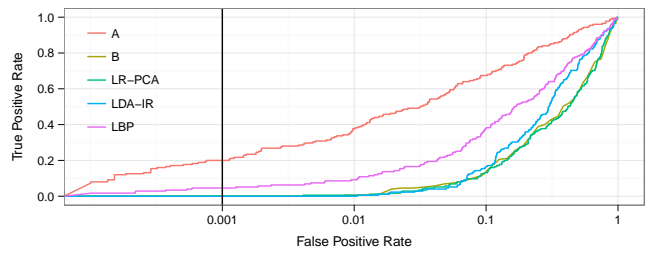
(a) Model Based Approach (with only morphable model texture)
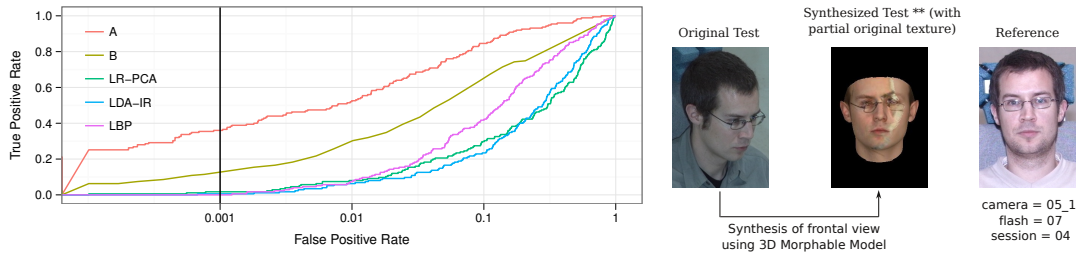
(b) Model Based Approach (with partial original texture supplemented by morphable model texture)

(c) View Based Approach (exact match of pose, illumination and camera)

(d) View Based Approach (large mismatch in pose and illumination)

Fig.	Test Set (or, Trace)	Suspect Reference Set	True Positive Rate (at FPR = 0.001)				
			A	B	LR-PCA	LDA-IR	LBP
(a)	*synth. frontal view (MM texture)	frontal view	0	0	0	0	0
(b)	**synth. frontal view (original tex.)	frontal view	0.36	0.13	0.01	0.01	0
(c)	surveillance view	surveillance view	**0.86**	**0.91**	0.23	0.20	**0.75**
(d)	surveillance view	near-surveillance view	0.2	0	0	0	0.05

(e) True Positive Rate corresponding to False Positive Rate of 0.001 for the model and view based approaches

Fig. 4: Face recognition performance using the model and view based approaches applied to a surveillance view test set. Note: A and B are commercial face recognition systems and the False Accept Rate axis is in log scale.