2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Person Recognition at a Distance: Improving Face Recognition Through Body Static Information

Ester Gonzalez-Sosa, Ruben Vera-Rodriguez, Javier Hernandez-Ortega and Julian Fierrez
School of Engineering, Universidad Autonoma de Madrid, Spain
ester.gonzalez@nokia-bell-labs.com
{ruben.vera, javier.hernandezo, julian.fierrez}@uam.es

***Abstract*—In this paper we evaluate body static information to improve the performance of face recognition at a distance. To this aim, we assess one state-of-the-art face recognition system based on deep features and three body-based person recognition systems, namely: i) row profiles with correlation coefficient, ii) row and column profiles with Support Vector Machines, and iii) contour coordinates with Dynamic Time Warping. Results are reported using the Multi-Biometric Tunnel Database, emphasizing on three distance settings: far, medium, and close, ranging from full body exposure to head and shoulders exposure. Several conclusions can be drawn from this work: a) row and column profiles are more robust than contour coordinates, b) face-based systems perform poorly at far distances, being body-based information more reliable at that distances, c) in general face-based systems perform better than body-based approaches at medium and close distances, and d) the multimodal fusion approach manages to outperform face-only recognition at distance in all distance-settings considered.**

I. Introduction

Is face recognition a solved problem? According to the Labeled Faces in the Wild face recognition benchmark (LFW), in which latest results are almost perfect, one may think that this problem is already solved [1]. However, results achieved in less constrained scenarios such as the recent initiatives *Megaface* [2] or *International Competition on Face Recognition in the Wild* [3] show that there is still room for improvement, especially concerning robustness towards aging, large scale, camera-subject distance and low resolution. These facts together with other challenging conditions encountered in surveillance and related unconstrained scenarios such as changes in pose, expression, illumination, blur or occlusions greatly hinder the performance of face recognition systems in the wild [4].

Aiming to enhance biometric system performance in challenging conditions, researchers have proposed multimodal approaches [5], that is, the use of additional biometric traits to complement or alleviate limitations given by a single biometric system. One useful case of multimodal approaches in unconstrained scenarios is the use of soft biometrics to complement the evidence given by face verification systems [6], [7], [8]. There exist several definitions of soft biometrics like the ones stated in [9], but most of them share the same main point: soft biometrics are based on information that can help to recognize people as it is related with each individual, but it is not enough to perform recognition accurately by themselves (see a review on soft biometrics in [9]).

However, when the person of interest is far from the acquisition system such it is the case of surveillance scenarios, it is not straightforward to estimate soft biometrics reliably, mainly due to the fact that the majority of the estimation algorithms are based on facial information [9], [10]. For those types of scenarios, Tome *et al.* proposed to use both face and body measurements to enhance face verification performance [7]. However, although the reported results by Tome *et al.* were promising, their measurements were manually annotated and therefore it is difficult to infer the real impact of soft information over face verification systems.

The main contribution of this paper is the proposal of a multimodal person recognition system that uses jointly face and body static information. The body-based person recognition is focused on shape information directly extracted from binarized images. Face recognition is based on a state-of-the-art deep learning approach. Performance of the individual systems along with the multimodal system are reported and discussed for three different distance settings within the Multi-Biometric Tunnel Database [11], varying from the far distance setting in which there is full-body exposure to the close distance setting in which only head and upper torso are visible. It is worth noting that our multimodal approach: i) extracts body and face information from the same single shot-image, contrary to other multimodal approaches based on face and gait [12], and ii) the whole multimodal approach is fully automatic, starting from face detection and background subtraction modules to the final multimodal decision.

The rest of this paper is organized as follows: Section II summarizes related works regarding body shape-based person recognition and multimodal approaches involving body information. Section III presents the proposed multimodal person recognition system. Then, Section IV describes the database and experimental protocol employed. Later, Section V presents the results. Finally, Section VI summarizes the contributions of this work and discuses future works.

II. Related Works

A. Body-based Biometrics

In this Section, some related works regarding the use of body information for person recognition purposes are reviewed. Although gait has been the mainstream approach for performing person recognition using body information [13], mainly based on body dynamic information from lateral

978-1-5386-3788-3/18/$31.00 ©2018 IEEE

views, here we summarize exclusively related works using information extracted from static images.

First of all, the use of body information has been widely used in the literature for people re-identification purposes, aiming to answer the question *"Where have I seen this person before?"* using visual cues such as color, texture or shape information [14], [15].

In what concerns person recognition, Nakajima *et al.* [16] explored color and shape-based features extracted from full body images. Results were reported using SVM as classifier and a small dataset composed of images from 8 different people with four different poses in different days. Very high performance (above 95% of accuracy) was observed when training and testing images were extracted the same day but the performance dropped significantly when the test set was not acquired the same day as the training images.

Hahnel *et al.* [17] further explored color and texture features with a database of 53 people, reaching recognition results of 97.4%. Authors acknowledged the limitations of using color information for some particular applications such as surveillance in shopping malls.

B. Multimodal Biometrics including Body Information

Multimodal approaches have been also proposed for person recognition in the literature to circumvent limitations of unimodal approaches in challenging scenarios. Information can be extracted from single or multiple sensors and fusion schemes can be applied at different levels [5]. In particular, fusion at score level is a popular option because of its simplicity and the performance increase that can be experienced. An extensive review of multimodal approaches for person recognition is beyond the scope of this Section, for a wide review see Ross *et al.* [5]. We restrict here to highlighting related works that have used dynamic or static body information as one of the biometric traits involved in the multimodal approach.

Concerning the use of body static information, Collins *et al.* [18] proposed to use body shape information like body height, width and some body part proportions to enhance performance of gait-based biometric systems. Shape and gait information were extracted from the same lateral views.

Tome *et al.* [7] proposed the use of face and body measurements in surveillance scenarios to enhance face verification systems. Concretely, this system was based on sparse representations and body measurements involving categorical information regarding both constitution and appearance information. Results were reported using 100 individuals from the Multi-Biometric Tunnel Database. Fusion led to improved verification results when introducing body information from 19% to 10% of EER and from 41% to 16% of EER for close and far distance scenarios, respectively. However, it is important to notice that body information was extracted from manual annotations.

III. Proposed System

In what follows we describe the general scheme of the proposed multimodal person recognition system designed for verification mode. As sketched in Fig. 1, the multimodal person recognition approach is divided into two branches, being the upper branch the face-based person recognition system and the lower branch the body-based person recognition system. On the upper branch, given a single-shot image, the face is automatically detected and preprocessed to further extract some discriminative features. Finally, a match score is estimated by comparing the extracted features with the template associated to the claimed identity. Likewise, on the lower branch, first the body silhouette is extracted through background subtraction techniques. Once the person is segmented, shape features are extracted from the binarized image. After comparing the extracted body shape test features with the body shape template associated to the claimed identity, the body based person recognition system outputs a match score. Finally, the multimodal fusion approach is built by fusing individual scores from face and body based person recognition systems.

A. Face-based Person Recognition System

The face-based person recognition system is divided into four stages, namely: face detection, preprocessing, feature extraction and matching. We proceed now to describe the details concerning each stage.

1) Detection and Preprocessing: Firstly, faces are detected through the Viola-Jones algorithm as it is known to perform reasonably well with frontal faces [19]. There are three possible outputs from the face detector: 1) One face is detected and the detector returns a bounding box that contains the face location. 2) Multiple faces are detected and the bounding boxes of all them are returned. Sometimes, there are false positive face detections and bounding boxes without a real face inside them are returned. In these cases, we decided to keep only the larger bounding box because in the majority of cases it contains the real face. 3) No face is detected so there is no output. This case is called Failed To Acquire (FTA) and happens frequently with images acquired at the far distance setting because of the low resolution. The FTA rate at far distance of this specific work is 20.41%. In these situations, the face-based verification system will not be able to provide any estimation. Once the face is detected, the image is histogram equalized using [20].

2) Feature Extraction: Face features considered in this work are based on deep learning. Concretely, features are extracted from pretrained CNN models through transfer learning techniques. In particular, two different pretrained models are assessed: Alexnet [21] and VGG-face [22]. We have performed some preliminary tests to decide which network to use. These tests have confirmed the intuition that VGG-face suits better to our requirements as it has been trained for face classification while AlexNet has been trained for object classification. Concretely, VGG-face was inspired by the previous VGG-Very-Deep-16 CNN network [23], using a dataset of 2.6 million faces and 2622 classes (individuals). Their deep architecture comprises 39 layers and it contains more than 130 million parameters.

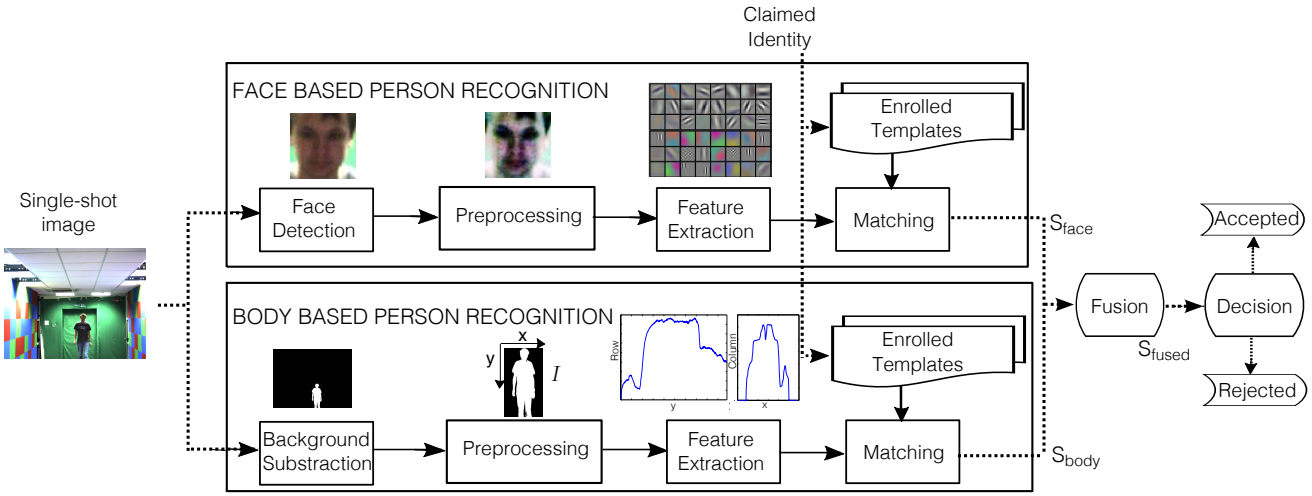


Fig. 1. **Multimodal person verification system**. Given a single shot image, both face and body discriminative features are extracted to further obtain individual scores, which are combined before the final decision. The figure visually shows the appearance of the learnable deep parameters and the row and column profiles involved in face and body-based modules, respectively.

In order to extract deep features from VGG-face, first we need to resize the images to the input size of the network, which is 224×224. Features are obtained by feedforwarding the images until the $fc7$ layer, which turned out to be the layer that achieves better verification results (compared to the $fc6$ layer). This way, for each image we obtain a feature vector of 4096 elements.

3) Matching: The last block of the face recognition system is the matching stage. As matcher we considered Support Vector Machines.

a) Support Vector Machines (supervised algorithm): SVMs is based on a representation of the examples as points in a multidimensional space. The training process consists in choosing a hyperplane that maximizes the distance from it to the nearest data point of each class. With this type of classifiers, it is interesting to have the input data represented in a high-dimensional space, what gives more freedom to find a hyperplane with a minimum distance large enough to obtain high classification rates.

B. Body-based Person Recognition System

This section describes a person recognition system based on body static information that can be used standalone or employed in a fusion scheme with other systems, for example with the face-based system from Section III-A.

1) Image Segmentation: In a real scenario, it would be essential to include a pedestrian detection module before extracting body information [24]. Since images from the Multi-Biometric Tunnel Database only contain one person per frame, the region of interest in which the person is located is inferred through a baseline background subtraction algorithm, setting the background as the initial frame of the sequence. The background subtraction is carried out in each of the 3 RGB channels and then the RGB result is grayscaled. The binarized silhouette is obtained after thresholding the gray scale image with a global threshold. The foreground segmentation is followed by some morphological noise reduction operations, in order to delete isolated foreground areas in the image. Finally, we proceed to select the bounding box around the foreground silhouette and normalize it to a height of 600 pixels and a different width for each distance scenario, according to the maximum width of the silhouette among all users for that scenario.

2) Feature Extraction: Two different shape feature extractors are considered, specifically: Row and Column Profiles (RCP) and Contour Coordinates (CC).

a) Row and Column Profiles: Given the binarized image I, pixels belonging to the foreground (fg) are set to 1 ($I(\mathbf{x}_{fg}, \mathbf{y}_{fg}) = 1$) and pixels belonging to the background (bg) are set to 0 ($I(\mathbf{x}_{bg}, \mathbf{y}_{bg}) = 0$). We then compute row and column profiles by counting foreground pixels across rows and columns, respectively.

b) Contour Coordinates: are also extracted from the normalized image I, being defined as $\mathbf{contour\_coordinates}(n) = (x_n, y_n), n = 1, ...n_{cc}$, being n_{cc} the number of pixels that compose the contour that the silhouette edge describes with the background, and (x_n, y_n) the coordinates of each one of those pixels. It is worth noting that each silhouette will have a different number of contour points, so the size of the feature vector is different between images. Unlike RCP, Contour Coordinates directly account for the relative location of the different points within the contour.

3) Matching: Two different distance-based matchers and one supervised classifier are explored, concretely: i) Dynamic Time Warping; ii) Correlation Coefficient, and iii) Support Vector Machines.

a) Dynamic Time Warping (unsupervised algorithm):: its goal is to find an elastic match among samples from two different sequences that minimize a given distance measure.

In this work, DTW is used to obtain the minimal cumulative distance between two sequences of contour coordinates, which do not have to share the same dimensionality. The algorithm searches for a path that minimizes the distance using a sequential procedure and holding some global and local constraints.

b) Correlation Coefficient (unsupervised algorithm):: the second classifier is based on the correlation of two feature vectors. The correlation coefficient measures the dependence between them, for example, if there exists a linear relationship. This measure of similarity is useful to compare row profile features as the persons to be identified will have the same distribution in the different parts of their body as opposed to row profiles pertaining to other identities. Column profile and contour coordinates can experiment more changes in their values as the person can change the relative position of its arms and legs, modifying significantly the feature vectors, so these features are not considered in this matcher.

c) Support Vector Machines (supervised algorithm):: this classifier has been used only with RCP as vectors to be compared must be of equal size (see Section III-A3).

C. Score Fusion

The final module of the multimodal person recognition system fuses individual scores. This fusion is carried out at score level. Previously, scores are normalized to the range $[0, 1]$ using the tanh-estimators described in [25].

The fusion method used consists of combining scores of both systems following the next logic [26], [27]. $a)$ If there are scores from both systems, then they will be added following the sum rule. $b)$ If one of the two systems does not have a score (e.g. due to face detection error), the resulting score will be a scaled version of the available score. $c)$ If there are no scores available, then the system will not be able to take a decision and that situation will be treated as a FTA.

The fusion scheme considered consists on: a weighted sum in which different weights are assigned to the different modalities at each distance-scenario, according to a specific criterion: $s_{fused} = p \cdot s_{face} + (1 - p) \cdot s_{body}$.

IV. Database and Experimental Protocol

A. Database

The dataset used for this work is a subset from the Southampton Multi-Biometric Tunnel database [11]. This database contains images of 227 different individuals walking through a tunnel in semi-unconstrained conditions and different distances. There are 10 sessions from each user, with each session having all the aforementioned information.

In this work we have considered only the frontal videos from the database to combine information from face and body at different distances. Similar to [7] we have defined three different scenarios varying the distance between the camera and the user (see Figure 5 in [7] for examples).

- **Far scenario**: images taken at 7.5 meters. Face is in low resolution (average resolution of 76×76), but the full body is available.
- **Medium scenario:** images taken at 4.5 meters. Face resolution is better than in the far scenario (average resolution of 135×135), but only the upper half of the body is visible.
- **Close scenario:** images taken at 1.5 meters. Face is in higher resolution (average resolution of 315×315), but only head and upper torso are available.

B. Experimental Protocol

The subset of the Multi-Biometric Tunnel Database considered in this work is comprised of images from the 227 subjects from 10 different sessions and at the 3 distance scenarios described before, resulting in a subset of $227 \times 10 \times 3$=6810 images. Results are reported for the verification mode in terms of ROC curves, Equal Error Rate and Verification Rate (EER and VR in %).

This whole set of images is divided into development and test sets following a 5-fold cross validation protocol, meaning that each time 4 folds out of 5 are used for development and the remainder one for test. The k individual results will then be averaged to produce a single performance value through their mean and standard deviation values. The experiments performed with development and test subsets are the same, being the purpose of the development dataset to calculate the statistics for the score normalization, so the process will be explained only for one subset.

In particular, each k-fold is further divided into train and test data. This time, a leave-one-out strategy is followed. For all the 10 sessions of each user we take 1 as the testing vector, and the rest (9) for training the classifiers. Concerning distance-based matchers, namely: DTW and Correlation Coefficient, the test feature vector is compared to the 9 training feature vectors yielding 9 individual scores. Then, these individual scores are averaged to obtain the final match score. Unlike distance-based matchers, the SVM classifier follows a slightly different protocol. For each user in the selected subset, a SVM model is trained using the 9 training feature vectors from that user as the positive class, and the training data from the remaining users as the negative class. When comparing the test feature vector to the different SVM models, a single score is directly obtained. We have decided to use a polynomial kernel of third grade, as it gives us the best results comparing it with other kernels and also trying to avoid overfitting without using polynomial kernels of higher grades.

V. Results

A. Individual Systems

Individual results for the face-based and body-based systems are summarized in Table I. As expected, the performance of the face-based system decreases as distance increases since less faces are detected and the discriminative information extracted from them is more limited due to low resolution.

In what concerns body-based system performance using row column profiles as features, it is observed that results improve with an increasing distance as there is more body information available in the images, but the best results are obtained at

EER[%]	**Far**	**Medium**	**Close**
Face System			
VGG-SVM	18.93	0.9	0.1
Body Systems			
CC-DTW	20.47	11.37	12.56
RP-CORR	15.01	12.19	19.57
RCP-SVM	**4.13**	**1.67**	**6.78**

TABLE I
EER OF THE INDIVIDUAL SYSTEMS OBTAINED FOR TEST DATA. RESULTS HAVE BEEN OBTAINED FOR FAR, MEDIUM AND CLOSE SCENARIOS, SIMILAR TO [7]. VALUES ARE EXPRESSED IN %. HIGHLIGHTED IN BOLD ARE THE BEST RESULTS FOR THE BODY INFORMATION SYSTEMS AT EACH DISTANCE.

medium distance, not at far distance. This fact shows that the upper body (head and torso) has more information about the user than the lower part. At medium distance, the upper torso is captured with higher resolution helping the classifier to discern better between users. Contour Coordinates also behave better at medium and close distance as these are the scenarios in which the contour contains more discriminative information among subjects. In terms of performance, the RCP + SVM body-based approach outperforms RP + CORR and CC + DTW approaches. It is interesting to note that the best performing approaches are SVM in both face-based and body-based systems. SVM are a more sophisticated classifier technology than the other alternatives, so their better results are not surprising. By analyzing individual performances from face-based and body-based recognition systems we observe that both systems could be complementary as the performance of one rises in the scenarios where the other worsens. Even more, at the far scenario, the body-based alternative obtains significantly better results than the face-based. These results encourage us to perform a fusion of both systems.

B. Score Fusion

The performance of the baseline fused system is reported in Table II. As can be seen, fusion results outperform the face-based systems at far distance and also body-based results at close distance. This makes sense because in each case, there is a system working properly and complementing the other. The criterion followed to estimate the weighting factor p has been set to the case in which at FAR=10^{-3} the verification rate (VR) is maximized. The experiments carried out for this estimation are depicted in Fig. 2. At far distance, the best verification rates are achieved with a low value for p, giving more importance to the body-based system. On the other hand, at close distance it happens the opposite, as best results are obtained giving more weight to the face-based system. At medium distance, in most systems, a halfway weight gives the best results.

The most important conclusion that can be extracted from these results is the high improvement of the global performance at far and medium scenarios compared to only using face recognition. These are the scenarios where the standalone face-based system fails more, but with the contribution of the body-based information a high number of those errors can be alleviated. Concretely, we infer that at far distance, the fusion scheme always improves the baseline system, while at medium and close distance it achieves it most of the times. These results demonstrate the utility of this multimodal biometric scheme in scenarios where the unimodal face-based system does not work optimally.

VI. CONCLUSION

In this work, we proposed a fully automatic multimodal person recognition system to enhance the performance of face recognition in challenging scenarios affected among other factors by low resolution and distance. We analyze the performance of individual systems (state-of-the-art face recognition based on deep learning and various body-based matchers) and their fusion in three different distance settings.

We have observed that: i) the face-based system performance increased with a decreased distance, performing very poorly in the far distance setting, and ii) the body-based system performs better at larger distances than close distances, noticing that upper body information is more discriminative than lower body information.

Regarding fusion schemes, we have learnt that: i) body shape information may be the best source of information to perform recognition when face is not detected or is poorly detected, ii) face-based performance is enhanced by body shape information specially in long distance scenarios where low resolution faces limit the face discrimination capability, and iii) person recognition performance is enhanced in medium and close distance settings when using RP-CORR or CC-DTW body shape systems.

Finally, some future work in this research line is discussed. Introducing automatic distance detection could be worth exploring to automatically configure fusion weights, together with more sophisticated quality-based biometric fusion schemes [27], [28] or pose-based score level fusion schemes [29]. Also testing this approach under more challenging databases such as the Point-and-Shoot Challenge [30] could provide insight regarding the real impact of this approach and unveil new challenges regarding people detection and segmentation. For that purpose it would also be essential to use state-of-the-art face detection algorithms [31].

ACKNOWLEDGMENT

This work was supported in part by Accenture, the project CogniMetrics from MINECO/FEDER under Grant TEC2015-70627-R, and the COST Action CA16101 (Multi-Foresee). The work of E. Gonzalez-Sosa was supported by a Ph.D. Scholarship from the Universidad Autonoma de Madrid.

REFERENCES

[1] G. B. Huang and E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep*, pp. 14–003, 2014.

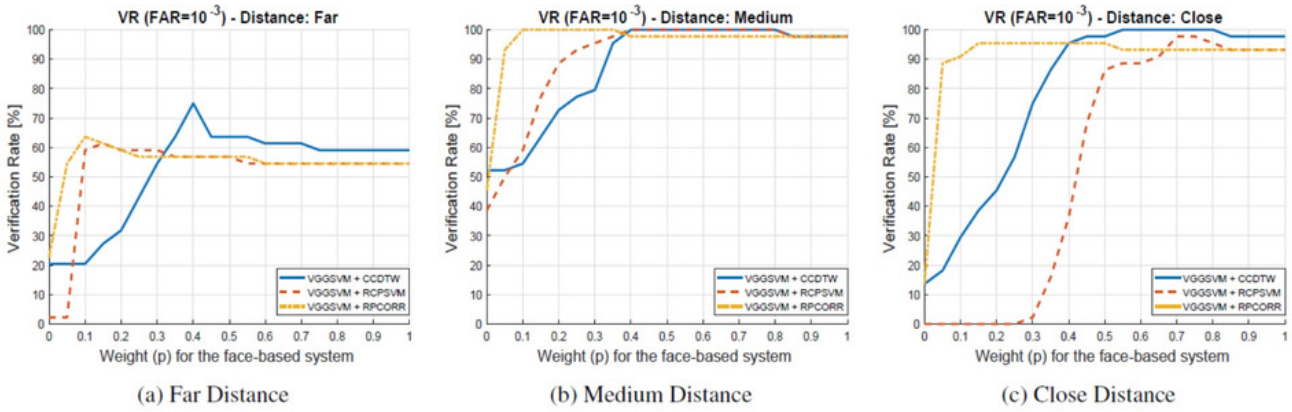
[2] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *Proc. of IEEE Proc. of CVPR*, 2016, pp. 4873–4882.

[3] J. Neves and H. Proença, "ICB-RW 2016: International challenge on biometric recognition in the wild," in *Proc. of IEEE Proc. of ICB*, 2016, pp. 1–6.

	Far		Medium		Close	
	VR (p opt.)	EER (p opt.)	VR (p opt.)	EER (p opt.)	VR (p opt.)	EER (p opt.)
Systems						
VGGSVM + RPCORR	63.64 (0.1)	11.36	**100 (0.35)**	0	95.45 (0.5)	0.13
VGGSVM + RCPSVM	61.36 (0.15)	4.86	**100 (0.4)**	0.053	97.73 (0.7)	0.34
VGGSVM + CCDTW	**75 (0.4)**	9.09	**100 (0.4)**	0.053	**100 (0.7)**	0.05

TABLE II

VR AND EER OF THE FUSION SYSTEMS. RESULTS ARE SHOWN FOR THE SCORE SUM WEIGHTS (p) THAT MAXIMIZE THE VR AT EACH SCENARIO WITH $s_{fused} = p \cdot s_{face} + (1-p) \cdot s_{body}$. RESULTS ARE SHOWN AS %. BETWEEN PARENTHESES IS THE VALUE OF p THAT MAXIMIZES THE VR IN EACH CASE. HIGHLIGHTED IN BOLD ARE THE BEST VR RESULTS AT EACH DISTANCE.



(a) Far Distance (b) Medium Distance (c) Close Distance

Fig. 2. **VR in function of sum weights.** FAR has been fixed to 10^{-3}. p is the sum weight for the face-based system and $(1-p)$ is the weight for the body-based scheme. The optimal weights change at different distance scenarios, as the performance of the individual systems that compose the fusion also varies.

[4] P. Dhar and A. Alavi, "Analysis of adaptability of deep features for verifying blurred and cross-resolution images," in *Proc. of IEEE Proc. of ISBA*, 2017, pp. 1–6.

[5] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics*. Springer, 2006.

[6] E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, "Facial soft biometrics for recognition in the wild: Recent works, annotation and cots evaluation," *IEEE TIFS*, vol. 13, no. 7, 2018.

[7] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE TIFS*, vol. 9, no. 3, pp. 464–475, 2014.

[8] H. Zhang, J. R. Beveridge, B. A. Draper, and P. J. Phillips, "On the effectiveness of soft biometrics for increasing face verification rates," *Computer Vision and Image Understanding*, vol. 137, pp. 50–62, 2015.

[9] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? a survey on soft biometrics," *IEEE TIFS*, vol. 11, no. 3, pp. 441–467, March 2016.

[10] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli, "On soft biometrics," *Pattern Recognition Letters*, vol. 68, pp. 218–230, 2015.

[11] R. D. Seely, S. Samangooei, M. Lee, J. N. Carter, and M. S. Nixon, "The University of Southampton Multi-Biometric Tunnel and introducing a novel 3d gait dataset," in *Proc. of BTAS*, 2008, pp. 1–6.

[12] E. Hossain and G. Chetty, "Multimodal identity verification based on learning face and gait cues," in *Proc. of Int. Conf. on Neural Information Processing*. Springer, 2011, pp. 1–8.

[13] M. S. Nixon, T. Tan, and R. Chellappa, *Human identification based on gait*. Springer, 2010.

[14] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Computing Surveys*, vol. 46, no. 2, p. 29, 2013.

[15] H. Jin, X. Wang, S. Liao, and S. Z. Li, "Deep person re-identification with improved embedding," *Proc. of IJCB*, 2017.

[16] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio, "Full-body person recognition system," *Pattern Recognition*, vol. 36, no. 9, pp. 1997–2006, 2003.

[17] M. Hahnel, D. Klunder, and K.-F. Kraiss, "Color and texture features for person recognition," in *Proc. of IEEE Int. Conf. on Neural Networks*, vol. 1, 2004, pp. 647–652.

[18] R. T. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Proc. of FG*, 2002, pp. 366–371.

[19] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[20] V. Struc *et al.*, "The INface Toolbox v2.0 The Matlab Toolbox for Illumination Invariant Face Recognition."

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[22] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *Proc. of Conf. on British Machine Vision*, 2015.

[23] A. Mahendran and A. Vedaldi, "Visualizing deep convolutional neural networks using natural pre-images," *International Journal of Computer Vision*, vol. 120, no. 3, pp. 233–255, 2016.

[24] A. Garcia-Martin and J. M. Martinez, "People detection in surveillance: classification and evaluation," *IET Computer Vision*, vol. 9, no. 5, pp. 779–788, 2015.

[25] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[26] P. Tome, J. Fierrez, F. Alonso-Fernandez, and J. Ortega-Garcia, "Scenario-based score fusion for face recognition at a distance," in *Proc. CVPR Workshop*, 2010, pp. 67–73.

[27] F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "Quality measures in biometric systems," *IEEE Security & Privacy*, vol. 10, no. 9, pp. 52–62, December 2012.

[28] J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho, "Multiple classifiers in biometrics. part 2: Trends and challenges," *Information Fusion*, vol. 44, pp. 103–112, November 2018.

[29] R. Kawai, Y. Makihara, C. Hua, H. Iwama, and Y. Yagi, "Person re-identification using view-dependent score-level fusion of gait and color features," in *Proc. of ICPR*. IEEE, 2012, pp. 2694–2697.

[30] J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer *et al.*, "The challenge of face recognition from digital point-and-shoot cameras," in *Proc. of BTAS*, 2013, pp. 1–8.

[31] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE TPAMI*, 2017.