

Expression-Invariant Multispectral Face Recognition: You Can Smile Now!

Ioannis A. Kakadiaris^a, George Passalis^a, George Toderici^a, Yunliang Lu^a,
Nikos Karampatziakis^a, Najam Murtuza^a, Theoharis Theoharis^a

^a Computational Biomedicine Lab, Dept. of Computer Science, Univ. of Houston, TX, USA

ABSTRACT

Face recognition performance has always been affected by the different facial expressions a subject may attain. In this paper, we present an extension to the UR3D face recognition algorithm, which enables us to diminish the discrepancy in its performance for datasets from subjects with and without a neutral facial expression, by up to 50%.

Keywords: Face Recognition, Multispectral Biometrics, Biometrics, Verification

1. INTRODUCTION

Information from the human face is one of the most natural, least intrusive biometrics. Achieving high verification rates in biometrics-based authentication/identification systems is essential for providing quicker airport check-ins, more secure paper-less banking, and many other applications, which require a person to confirm their identity. Facial expressions, such as smiling, have always undermined the performance of face recognition systems to such a degree that certain countries now require the subjects applying for passports to pose with a neutral expression.

We have proposed a method which achieves high verification rates even in the presence of facial expressions. Specifically, we use images from multiple cameras to reconstruct the 3D face geometry of the subject, on which we map the data from a co-registered infrared camera. Then, we fit an annotated face model with imposes an explicit two-dimensional parameterization to this data. Using the fitted model, we construct a three-channel UV deformation image encoding the facial geometry, and a single channel UV vasculature map obtained from the infrared (IR) image. We use these maps, and additionally the visible texture, for computing the similarity between the subject and entries in a database. The novelty of our work lies in the use of deformation images and physiological information as means for comparison.

In this paper, we propose a new similarity metric which increases the performance and robustness of our system with respect to facial expressions. Specifically, we performed extensive tests on the Face Recognition Grand Challenge¹ (FRGC) dataset v2, and on our own multispectral dataset. Our algorithm achieves the best results on both datasets in the presence of facial expressions as compared with other published algorithms (to the best of our knowledge) when the 3D data are used only.

The rest of the paper is organized as follows. Section 2 reviews related work, while Section 3 details our new similarity metric. Section 4 presents our results and their implications.

2. PREVIOUS WORK

Due to limited space, we present only recent work in the area. Comprehensive review of the previous work can be found at.² Milan *et al.*³ uses a tensor representation for 3D faces, and a similarity measure based on the linear correlation coefficient. However, the reported results indicate only 86.4% rank-1 recognition on a database of 277 subjects (a small subset of the FRGC v2 dataset). Recently, Viisage⁴ reported a hierarchical graph matching technique that achieves verification rate of 96.8% at 0.1% FAR (ROC III) of the FRGC dataset when fusing the results of the analysis of both the 2D and 3D data. However, they report verification rate of 89.5% at 0.1% FAR (ROC III) when using the 3D data only.

Further author information: (Send correspondence to I.A.K.)
I.A.K.: E-mail: ioannisk@uh.edu, Telephone: 1 713 743 1255

3. METHODS

In this paper, we present an extension to our previous face recognition algorithm, consisting of an improved similarity metric. Here is a summary of our approach:

1. *Acquisition*: Acquire raw data in the form of frames from calibrated visible spectrum cameras and calibrated IR spectrum camera.
2. *Preprocessing*: Reconstruct a 3D mesh and a thermal texture map of the face from the raw data.
3. *Model Fitting*: Fit the AFM to the mesh, using the original procedure.²
4. *Parametric Deformation Image*: Use the UV parameterization of the fitted AFM to generate a three-channel deformation image of the face by encoding its geometry.
5. *Parametric Thermal Image*: Use the UV parameterization of the fitted AFM to generate a one-channel thermal image of the face by encoding its temperature.
6. *Skin Mask*: Segment the thermal image to construct a mask identifying the regions of UV space corresponding to skin and use this as a mask to weigh the deformation image.
7. *Metadata Extraction–Deformation Image*: Compress the deformation image using a wavelet transform using the procedure described in Section 3.1 in addition to the original Haar² decomposition.
8. *Metadata Extraction–Vasculature Image*: Extract the facial vasculature in the forehead.
9. *Storage/Comparison*: If in the enrollment phase, store the compressed deformation image, 2D still and vasculature image (if applicable) to the database along with additional information (e.g., name, age, date of enrollment). If in the identification/verification phase, compare probe data with the gallery data as described in Section 3.1.

3.1. Metadata Representation

Currently, we employ two similarity metrics: the L^1 metric applied on Haar wavelets,² and an extension based CW-SSIM,⁵ which we will present below. The scores generated by these two metrics are then fused to obtain optimum result.

Specifically, the deformation image is decomposed using the complex version⁶ of the steerable pyramid transform,⁷ a linear multi-scaled, multi-orientation image decomposition algorithm. The resultant wavelet representation is translation-invariant and rotational-invariant. This feature is desirable to address possible positional and rotational displacements caused by facial expressions. Our algorithm applies a 3-scale, 10-orientation complex steerable pyramid transform to decompose each channel of the deformation image. Only the low-pass orientation subbands at the furthest scale are stored. This enables us to compare two deformation images directly and robustly using multiple orientations.

In order to quantify the distance between the two compressed deformation images of the probe and gallery, we need to assign an numerical score to each part F_k of the face. Note that F_k may be distorted in different ways by facial expressions. To address this, we employ the complex wavelet structural similarity (CW-SSIM) index algorithm. CW-SSIM is a translational insensitive image similarity measure inspired by the structural similarity (SSIM) index algorithm.⁸ A window of size 3 is placed in the X direction (first channel). The window then traverses across the image one step at a time. In each step, we extract all wavelet coefficients associated with F_k . This results in two sets of coefficients $p_{w,i} = \{p_{w,i} | i = 1, \dots, N\}$ and $g_{w,i} = \{g_{w,i} | i = 1, \dots, N\}$, drawn from the probe image and the gallery image respectively. The distance measure between the two sets is a variation of the CW-SSIM index originally proposed by Wang⁵:

$$\tilde{S}(p_w, g_w) = 1 - \left(\frac{2 \sum_{i=1}^N |p_{w,i}| |g_{w,i}| + K}{\sum_{i=1}^N |p_{w,i}|^2 + \sum_{i=1}^N |g_{w,i}|^2 + K} \right) \cdot \left(\frac{2 |\sum_{i=1}^N p_{w,i} g_{w,i}^*| + K}{2 \sum_{i=1}^N |p_{w,i} g_{w,i}^*| + K} \right)^r$$

The first component measures the equivalence of the two coefficient sets. If $P(w, i) = G(w, i)$ for all i 's, the perfect distance 0 is achieved. The second component reflects the consistency of phase changes. Since facial expressions are present, addressing spatial distortion is more favorable than measuring equivalence. The exponent r is determined experimentally. As the sliding window moves, the local $\tilde{S}(p_w, g_w)$ at each step w is

Table 1. FRGC Baseline Verification Rate at 0.1% False Accept Rate.

Description	ROC1	ROC2	ROC3
Full Set	0.6720	0.6076	0.5342
Neutral Expressions	0.9000	0.8598	0.8158
Non-Neutral Expressions	0.3688	0.2598	0.1695

computed and stored. The weighted sum of the local scores from all windows gives the distance score of F_k in the X direction: $Score_x(F_k) = \sum_{w=1}^N (b_w \cdot \tilde{S}(p_w, g_w))$ where b_w is a predefined weight depending on which subband the local window lies on. The weight assigned to a particular subband is determined by the descriptive power and performance, which varies dramatically due to the subband’s distinct orientation.

Similarly, we compute the scores for the Y and Z directions. By summing the scores in the X, Y and Z directions, the total distance score of F_k is computed. The discrete weighted sum of the scores for all F_k ’s is the overall distance between the probe image P and the gallery image G :

$$Distance(P, G) = \sum_{k=1}^N (w_k \cdot (Score_x(F_k) + Score_y(F_k) + Score_z(F_k)))$$

where w_k is a predetermined weight for area F_k . The distance between a probe and a gallery is computed as the weighted sum of the CW-SSIM and L^1 metrics.

4. RESULTS & DISCUSSION

In order to assess the performance of the UR8D algorithm in the presence of facial expressions, we use the FRGC dataset. We chose to use the data for Experiment 3, which contains 4007 3D scans and their corresponding 2D texture images. The baseline algorithms for matching for 3D data are based on the work of Chang *et al.*,⁹ and for 2D data on the work of Moon and Phillips¹⁰ and the algorithms from the CSU Face Identification Evaluation System.¹¹ We will report scores using three Receiver Operator Characteristic curves (denoted as ROC1, ROC2, ROC3), as defined in FRGC. Moreover, we will report scores for the full set, and for subsets which correspond to each of the annotated facial expressions.

The ROC graphs for both the FRGC baseline and the UR8D algorithm can be found in Figure 1. We summarize the verification rates at 0.1% False Accept Rate (FAR) in Table 2 for UR8D, and in Table 1 for the FRGC baseline. The results show that the FRGC baseline is extremely sensitive to facial expressions, even though, unlike UR8D, it makes use of both texture and shape information. For the 0.1% FAR point, the verification rate drops by 79.22% in the case of ROC3.

In the presence of facial expressions, the maximum performance drop for UR8D occurs for ROC1, and it is only 5.36%. Table 2 shows the performance for the different annotated facial expressions found in the FRGC database. When comparing the difference in performance between neutral and non-neutral expressions, the UR8D algorithm has improved by 51.02% for ROC1, 52.76% for ROC2 and 54.09% for ROC3, with respect to the previous version.¹²

The majority of our failures were caused by exaggerated facial expressions in which the nose was shrunk or expanded. Other failures were caused by a combination of large movement in the eyebrow region, and lifted or puffy cheeks. In the failures, the facial expressions have caused a misalignment between the AFM and the data. If the alignment is not proper, then the fitting will generate a deformed model whose annotation is no longer consistent with the physical landmarks of the face. Therefore, when comparing the geometry image generated by such a fitting with a proper geometry image, the difference will always be high.

5. CONCLUSION

We have demonstrated the ability of our latest algorithm to handle various facial expressions without a significant performance loss, when compared to previous approaches.

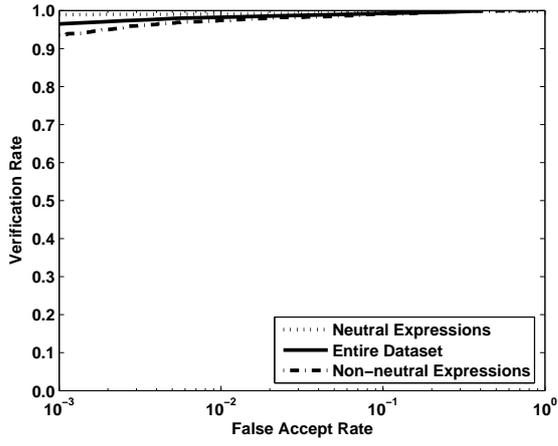
Table 2. UR8D Verification Rates for 0.1% False Accept Rate.

Description	ROC1	ROC2	ROC3
Full Set	0.9651	0.9621	0.9592
Neutral Expressions	0.9895	0.9874	0.9850
Non-Neutral Expressions	0.9365	0.9355	0.9352
Disgust	0.9704	0.9703	0.9706
Happiness	0.9673	0.9651	0.9629
Other	0.9722	0.9674	0.9627
Sadness	0.9651	0.9611	0.9568
Surprise	0.9652	0.9611	0.9574

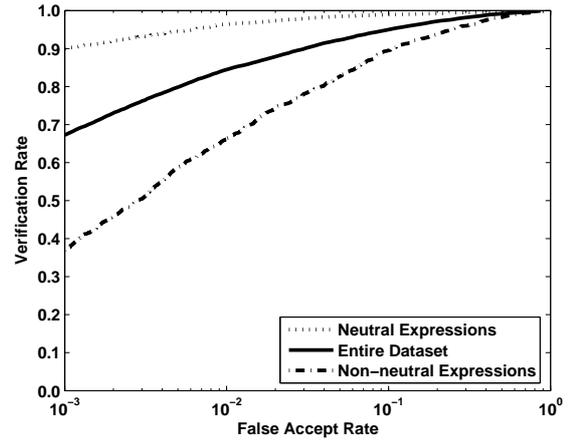
The most important improvement is the choice of distance metrics. Previously, we have used the L^1 norm applied to Haar-transformed geometry images. This metric performs very well when there are no facial expressions present. However, a more robust metric, the CW-SSIM, applied to steerable pyramid transforms of the geometry images is more robust with respect to local deformations which may be caused by the muscle movement in the subject’s face when having a pronounced facial expression. The fusion between of the scores yielded by the two metrics yielded our best result presented in this paper.

REFERENCES

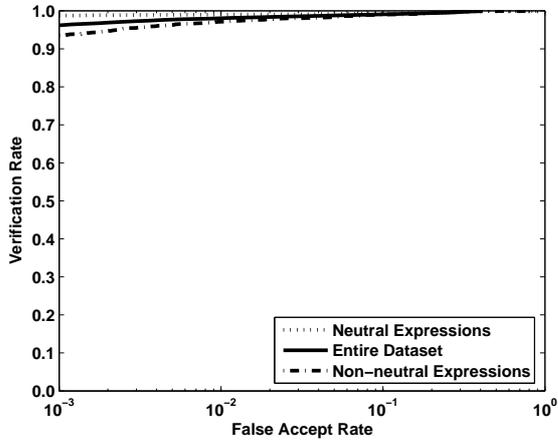
1. P. Phillips, P. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “Overview of the Face Recognition Grand Challenge,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, (San Diego, CA), June 2005.
2. I. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza, “Multimodal face recognition: Combination of geometry with physiological information,” in *Proc. IEEE Computer Vision and Pattern Recognition*, **vol. 2**, pp. 1022–1029, (San Diego, CA), June 20-25 2005.
3. A. Milan, M. Bennamoun, and R. Owens, “Matching tensors for pose invariant automatic 3d face recognition,” in *Proc. IEEE Workshop on Advanced 3D Imaging for Safety and Security*, (San Diego, CA), 2005.
4. S. G. M. Hüsken, M. Brauckmann and C. Malsburg, “Strategies and benefits of fusion of 2d and 3d face recognition,” technical report, Viisage Technology AG, 2005.
5. Z. Wang and E. P. Simoncelli, “Translation insensitive image similarity in complex wavelet domain,” *IEEE International Conference on Acoustics, Speech and Signal Processing* **vol. II**, pp. 573–576, March 2005.
6. J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistic of complex wavelet coefficients,” *Int’l J Computer Vision* **vol. 40**, pp. 49–71, 2000.
7. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multi-scale transforms,” *IEEE Trans Information Theory* **vol. 38**, pp. 587–607, 1992.
8. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans Image Processing* **vol. 13**, pp. 600–612, Apr 2004.
9. K. Chang, K. Bowyer, and P. Flynn, “Face recognition using 2D and 3D facial data,” in *Proc. Workshop on Multimodal User Authentication (MMUA)*, pp. 25–32, (Santa Barbara, CA), Dec. 2003.
10. H. Moon and P. J. Phillips, “Computational and performance aspects of PCA-based face-recognition algorithms,” *Perception* **30**(3), pp. 303–321, 2001.
11. R. Beveridge and B. Draper, “Evaluation of face recognition algorithms (release version 5.0).” [online] <http://www.cs.colostate.edu/evalfacerec/index.html>.
12. G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, “Evaluation of 3d face recognition in the presence of facial expressions: an annotated deformable model approach,” in *Proc. IEEE Workshop on Face Recognition Grand Challenge Experiments*, (San Diego, CA), June 20-25 2005.



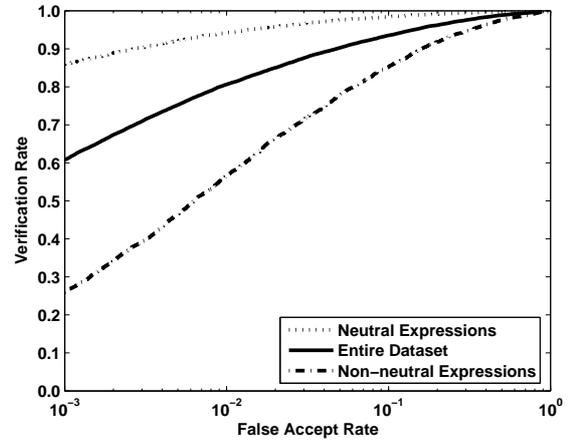
(a)



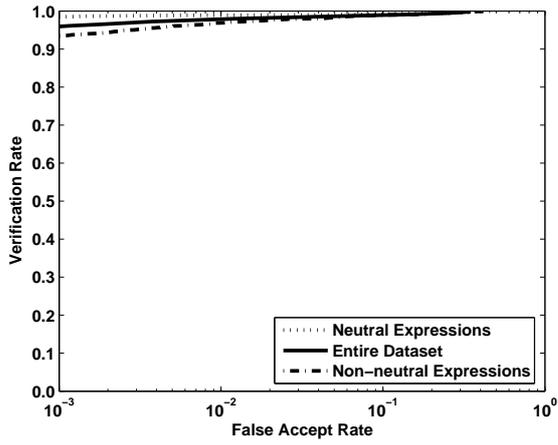
(b)



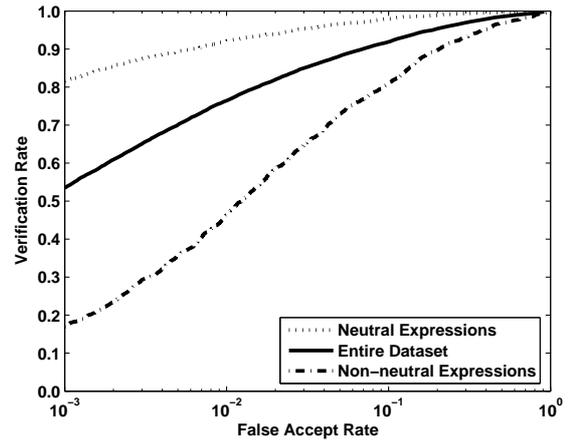
(c)



(d)



(e)



(f)

Figure 1. UR8D Performance: (a) ROC1, (c) ROC2, (e) ROC3; Baseline Performance: (b) ROC1, (d) ROC2, (f) ROC3