BIOMETRIC IDENTIFICATION USING FACIAL MOTION AMPLIFICATION

T. Pistola[1]*, A. Papadopoulos*[1]*, N. Mitianoudis*[1]*, and N. V. Boulgouris*[2]

[1]Democritus University of Thrace
Electrical & Computer Eng. Dept.
Xanthi 67100, Greece

[2]Brunel University London
Electronic & Computer Eng. Dept.
United Kingdom

ABSTRACT

We propose a new biometric trait based on facial motion amplification. The main advantage of the new biometric characteristic is that it does not rely on the visibility of critical facial features, such as nose, mouth, iris, or eyebrows. This makes it effective even when the respective areas are covered. Using the proposed system, facial image sequences are captured using an ordinary video camera and facial blood flow is calculated by means of small motion amplification. The calculated blood flow is captured from limited facial areas and is represented as a template that is suitable for identification purposes. Experiments on a new database show promising performance of the proposed approach, and provide evidence of the discriminatory capacity of the proposed biometric.

***Index Terms*—** Biometrics, Motion Amplification, Facial Blood Flow

1. INTRODUCTION

The importance of biometric identification has been increasing steadily in the past couple of decades and this trend is unlikely to diminish soon. Traditional biometric systems, such as those performing fingerprint, iris, or face recognition, have evolved significantly over the years and are now widely used in numerous identification and access control situations due to their excellent identification capabilities, even when applied to large databases.

The evolution of traditional biometric methods, however, has taken place in parallel with the invention of new biometric modalities, such as palm recognition, vein recognition, or ear recognition. Such new biometrics either aim to achieve more effective stand-alone identification or offer superior performance through their use in multi-biometric systems [1]. In this context, each new biometric approach complements existing biometrics and strengthens the performance of traditional biometric systems.

The advent of methods for the detection and analysis of micro movements [2] offers a new direction for devising and deploying new biometrics. In this paper, we propose the use of patterns representing Facial Blood Flow (FBF) as a new way to identify individuals. Specifically, we examine the discriminatory capacity of Facial Blood Flow (captured in a contactless way) and we try to reach conclusions regarding its potential use as biometric. Our proposed method does not directly assess the similarity of facial texture between subjects, but instead it extracts spatio-temporal blood flow information, which is used as a distinctive biometric. The proposed method is tested by observing image sequences depicting small facial areas, which exhibit variations in blood distribution during each heartbeat cycle. Although from the outset it can be claimed that such an approach will not be able to rival the current best-performing identification strategies, the proposed approach has a number of features that make it an excellent complement to traditional face recognition.

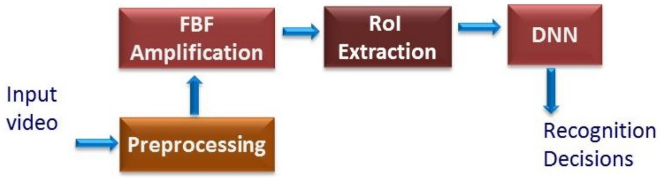

Fig. 1. Block diagram of the proposed person identification system based on Facial Blood Flow.

The paper is organized as follows. In Section 2, motion amplification is outlined as the basis for the extraction of the new biometric information. In Section 3, blood flows are extracted from specific facial areas and combined onto a single feature vector. In Section 4, classification is performed using a deep neural networks. Experiments are presented in Section 5, in order to show the discriminatory capacity of the new biometric. Finally, conclusions are drawn in Section 6.

2. MOTION AMPLIFICATION

Video Amplification was initially proposed by Wu et al. [2]. This seminal paper introduced a technique for amplifying small motions in common videos and make them visible to the human eye. As an initial example, the authors of [2] demonstrated that the non-visible flow of blood in the face

Fig. 2. Sample frames showing Facial Blood Flow extracted using motion amplification.

can be amplified and be observed by the human eye. In [2], each frame of the video sequence is decomposed into a Laplacian pyramid, thus constructing a 3D stack. The first two dimensions contain the Laplacian pyramid decomposition and the third dimension its evolution over time. Wu et al. [2] process the temporal evolution of each pixel separately in order to reveal the hidden motion that exists in the video. Additional mathematical insight into this procedure is presented in [2], [4].

The work in [3] presented extensions of the framework in [2]. The first extension was to use a complex steerable pyramid decomposition instead of the Laplacian pyramid decomposition. In order to improve the motion amplification of the image's edges, i.e., amplify the non-visible oscillations of "still" objects in the scene, they apply the same amplification procedure, as described earlier, but only on the phase of the complex steerable pyramid decomposition. This is based on the observation that the phase of the complex steerable decomposition, in a similar manner to the Fourier transform, contains most of the image's edge information.

In this paper, we use the methods in [2, 3] in order to capture motion information that can serve as the basis for the extraction of hidden human features with discriminatory capacity.

3. FACIAL BLOOD-FLOW BASED IDENTIFICATION

3.1. Video capture and pre-processing

An outline of the proposed system is shown in Fig. 1. The first step of our method is to capture the subject's face using an imaging device. Since we want to monitor blood flow, this implies a periodical phenomenon of no more than 2 Hz (60-90 pulses per min is the average heart rate for a resting person). Thus, a high frame-rate camera is not required for this experiment. An ordinary camera of 30 fps (i.e. 30 Hz) is sufficient to capture the facial blood flow, since the Nyquist frequency is 4 Hz. The second requirement is that the subject stays as still as possible during the capture. We need to have two captures of each subject. The first one will be used for the system's training, whereas the second will be used for the system's testing phase. We also used natural day light in the capture room, in order to avoid additional oscillations by artificial light (i.e. oscillations from the mains).

No human being can stand perfectly still for more than a few seconds, and even then, there are still small motions that can affect the result, especially after a large amplification. Given these conditions, the motion amplification will amplify these motions, thus increasing the noise of the final signal, or even destroying the result, because of very large motions. So, the next step in our system is to attenuate these larger motions, without much affecting the result. This is achieved by using motion amplification with a negative amplification factor α, in the range of [-1,0). This way, the motions not related to the blood movements are attenuated, but not removed [3]. In our approach, we used the MATLAB code provided by [3] for phase-based attenuation.

3.2. Facial Blood Flow Amplification

Once the large motions have been attenuated, the next step is to amplify the blood-flow in the face. This is performed using the Eulerian Video Magnification method by Wu et al. [2]. We used an amplification factor $\alpha = 350$ for this task. Through this process, we obtain a sequence of images $\mathcal{F}$ representing blood flow in the subject's face. To reduce the computational complexity, we formed a grayscale version of that image sequence, whereby pixel intensity is proportional to blood flow in the subject's face. The frequencies that are amplified by the system are between 1 and 2 Hz, which were chosen to match the average normal human heart rate. In this way, unwanted micro-movements are not amplified. Finally, since we are interested only in the blood flow and not in the actual video content, the output of this stage keeps only the amplification that is added to the original video. Some RGB frames from a typical amplified Facial Blood Flow video are shown in Fig. 2.

3.3. Extraction of facial regions of interest

It is computationally inefficient to process the whole amplified video extracted from the previous stage. In addition, several features of people's face may hinder the view of facial blood-flow. These features may include possible facial hair, such as a beard and a moustache. Moreover, other features, such as the eyes and consequently eye-flickering, may also confuse the recognizer. For this reason, we focus only on feature regions where blood flow is visible and easily identifiable. Thus, we agreed on using the areas of the two cheeks and the forehead. These areas are usually covered by human skin and possible hair on the forehead can be removed temporarily for identification.

To automatically locate these regions of interest, the first step is to detect the face area in the motion-attenuated video. This is performed using the algorithm of Zhu and Ramadan [5]. The next step is to identify several keypoints/landmarks in the human face. To achieve this, we fitted an Active Appearance Model (AAM) [6]. The AAM matches a set of standard points/landmarks on the face. This is performed for all video frames. In our study, we used the robust implementation of AAM fitting by Bulat and Tzimiropoulos [7]. Once we know the position of several face landmarks (see Fig. 4(b)), we can approximately define the position of three areas of interest. The first one is a lower part of the forehead. The other two are parts of the cheeks, one left and one right. To keep the algorithm simple, we defined these three regions as rectangles of fixed dimensions. In our experiments, the forehead region was selected to be a rectangle of dimensions 70×100, while the left and right cheek regions were selected to be rectangles of dimensions 50×50. The exact position of the forehead region is determined with reference to key-points on the eyebrows and the nose. The position of the cheek regions is determined by correlating key-points on the lower-eyes and nose. An example of the detected regions of interest is shown in Fig. 4(c).

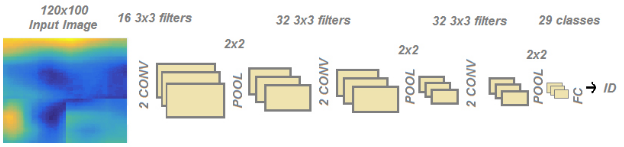

Fig. 3. Convolutional neural network for template classification.

3.4. Construction of Facial Blood Flow (FBF) templates

For recognizing, we define a template $\mathcal{T}$, which is constructed by averaging consecutive blood flow images $\mathcal{F}$ representing blood flow in a short period of time. In this work we considered sequences of 1 second, as this will allow us to capture at least one heart cycle. For region r, with $r = 1, 2, 3$, we define template $\mathcal{T}_r$ as

$$\mathcal{T}_r = \frac{1}{N} \sum_{t=1}^{N} \mathcal{F}_t^r(x, y) \tag{1}$$

where (x, y) denotes pixel spatial coordinates and $\mathcal{F}_r$ denotes the blood flow region r at time $t, t = 1, \ldots, N$. The number N of frames used for the construction of the template was chosen so as to correspond to 2 seconds. The 3 regions $\mathcal{T}_r$ are collated to form a single template image $\mathcal{T}$ (see Fig. 4(e)).

4. PERSON IDENTIFICATION USING FCNS AND CNNS

The template images (see Fig. 4(e)) constructed in the previous stage are input to a supervised classifier. In this work, we experiment with deep Fully Connected Neural Network (FCN) and Convolutional Neural Network (CNN) classifiers [8].

We examined three FCN and one CNN architecture. Although in all architectures the number of neurons in the input layer is equal to the number of pixels on the FBF template, the rest of these architectures differ. Specifically, FCN1 is a shallow architecture, consisting of 3 layers with 200, 100 and 29 neurons each. The last layer is the classification softmax layer with 29 outputs, equal to the number of classes (people) in the experiment. FCN2 is a deeper architecture consisting of 5 layers with 200, 200, 100, 50 and 29 neurons each. Finally, FCN3 consists of 6 layers with 500, 500, 500, 250, 100 and 29 neurons. All three networks feature a Dropout mechanism before the last layer with $p = 0.5$. All neurons use the ReLU activation function. The CNN architecture consists of 3 convolutional stages (16, 32 and 32 3×3 filters respectively) and 1 FCN stage (512 neurons). All convolutional stages feature the ReLU activation function, a 2×2 Max-Pooling stage and a dropout layer with $p = 0.25$. The FCN stage uses the ReLU and a dropout layer with $p = 0.5$. The last layer is a softmax classification layer.

For all architectures, we used the Adam optimizer [9] with the categorical cross-entropy loss function, since we deal with more than 2 output classes [8]. The learning rate was set to $\eta = 0.01$ and the networks were trained for 800 epochs.

5. EXPERIMENTAL EVALUATION

5.1. Dataset

To evaluate the effectiveness of Facial Blood Flow (FBF) as a biometric trait, we recorded a new dataset, under recording conditions that took into account possible problems arising during the subsequent motion amplification. For the recording of facial image sequences, we used a GoPro Hero 4 Black camera with 1280×720 resolution and 30 frames per second. A total of 29 subjects were recorded in a room with natural light, which helped avoid oscillations from artificial light sources. The subjects were seated on a chair at a fixed distance from the camera and were instructed to stay as motionless as possible during the recording. For each subject, we conducted two almost consecutive 20-second recordings, which were used for training purposes. For testing, we captured another 20-second recording from each subject after an hour. The compiled database is summarized in Table 1.

Table 1. Description of the new dataset.

Database	
Number of subjects	29
Frame-rate (fps)	30
Resolution	1280 X 720
Sequence duration (seconds)	20

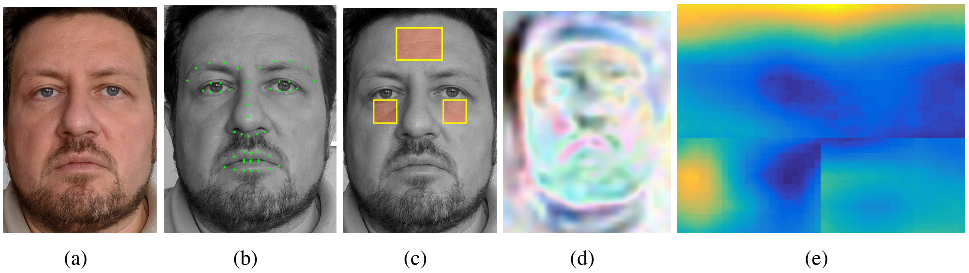

Fig. 4. (a) Original face, (b) AAM fit on face (c) Detection of the proposed facial regions of interest, (d) Blood flow representation for the entire face, constructed from motion-amplified face images in a period of 2 seconds, (e) Template $\mathcal{T}$ containing only the collated regions of interest (forehead and two cheeks, yellow-box regions).

5.2. Experiments

We implemented the Movement Motion Attenuation and the Facial Blood Flow Amplification stages based on the code provided by [2] and [3] respectively. The face isolation and AAM fitting stages were based on the method in [7]. We implemented the four DNN architectures in Python using TensorFlow, Keras and an NVidia Titan X Pascal GPU. The two systems were trained using the training dataset for 800 epochs. We consequently used the testing dataset for assessing the system's performance.

Since to the best of our knowledge there are no other works exploring the use of FBF as a biometric, we cannot conduct a straightforward comparison of our system to other systems. Therefore, we focus on investigating the inherent discriminatory capacity of FBF and its potential use as a biometric trait. To this end, we first measure the performance of our methodology using only the forehead patch. Subsequently, we use data from all three facial regions shown in 4, which do *not* include critical features, such as nose, mouth, iris, or eyebrows.

Results are summarized in Table 2. As seen, between the four architectures, the deepest CNN architecture (CNN) yields the best results, which implies that spatial information is important. Between the FCN architectures, the shallow FCN1 exhibits the best performance. Interestingly, identification performance is very promising even when only the forehead data are used. When all three areas are used, the best recognition rate achieved is 93.84%. These results demonstrate that Facial Blood Flow (FBF) has discriminatory capacity that can be used for person identification. Although FBF information from the forehead alone can act as a basic discriminator, the additional contribution from the other two facial regions yields significantly increased performance, which is impressive, considering that no facial texture was used and no conventional facial features were taken into account in the classification stage.

Considering that the proposed system is rather simple, it is expected that improvements in the system's various processing stages will result in reliable performance even in cases when smaller facial patches are used. This methodology can potentially lead to biometric systems that can operate reliably even when important parts of the face are covered. This will result in effective facial recognition technology that does not rely on the availability of critical facial data.

Table 2. Recognition Accuracy for the four classification architectures using forehead data only or using all three areas.

Architecture	Forehead	Three areas
FCN1	68.63 %	87.96 %
FCN2	72.55 %	84.59 %
FCN3	60.22 %	85.15 %
CNN	71.43 %	93.84 %

6. CONCLUSIONS

We proposed a new biometric trait based on facial motion amplification. Using the proposed system, facial image sequences were captured using an ordinary video camera and facial blood flow was calculated by means of small motion amplification. The calculated blood flow was captured from limited facial areas and was represented as a template that is suitable for identification purposes. Experiments on a new dataset showed the promising performance of the proposed system and provided evidence of the discriminatory capacity of the proposed biometric.

Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research. We would also like to thank Dr. Georgios Tzimiropoulos for sharing his Active Appearance Model fitting python code. We would also like to thank the 29 students and staff at the Electrical and Computer Eng. Dep. at Democritus University of Thrace, who participated in our experiment.

7. REFERENCES

[1] A. A. Ross, K. Nandakumar, A. K. Jain, Handbook of Multibiometrics, Springer, 2006.

[2] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, W.T. Freeman, “Eulerian Video Magnification for Revealing Subtle Changes in the World”, *ACM Trans. on Graphics*, Vol. 31, No. 4, 2012.

[3] N. Wadhwa, M. Rubinstein, F. Durand, W.T. Freeman, “Phase-based Video Motion Processing”, *ACM Trans. on Graphics*, Vol. 32, No. 4, 2013.

[4] M. Rubinstein, *Analysis and Visualization of Temporal Variations in Video*, PhD Thesis, Massachusetts Institute of Technology, 2013.

[5] X. Zhu, D. Ramanan, “Face detection, pose estimation and landmark localization in the wild”, *Computer Vision and Pattern Recognition (CVPR)*, Providence, Rhode Island, USA, 2012.

[6] T.F. Cootes, G.J. Edwards, C.J. Taylor, “Active Appearance Models”, *European Conference on Computer Vision (ECCV)*, Freiburg, Germany, 1998.

[7] A. Bulat, G. Tzimiropoulos, “Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources”, *IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, 2017.

[8] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.

[9] D.P. Kingma, J. Ba, “Adam: A Method for Stochastic Optimization”, *3rd Int. Conf. on Learning Representations (ICLR 2015)*, San Diego, CA, 2015.