Article

Person Identification using Temporal Analysis of Facial Blood Flow

Maria Raia [1], Thomas Stogiannopoulos [1], Nikolaos Mitianoudis[1,*] and Nikolaos V. Boulgouris [2]

[1] Democritus University of Thrace, University Campus Kimmeria, 67100 Xanthi, GR; {mariraia, tstogian, nmitiano}@ee.duth.gr
[2] Brunel University London, College of Engineering, Design, and Physical Sciences, Uxbridge, Middlesex UB8 3PH UK; nikolaos.boulgouris@brunel.ac.uk
* Correspondence: nmitiano@ee.duth.gr; Tel.: +302541079572

Abstract: Biometrics play an important role in modern access control and security systems. The need of novel biometrics to complement traditional biometrics has been at the forefront of research. The Facial Blood Flow (FBF) biometric trait, recently proposed by our team, is a spatio-temporal representation of facial blood flow, constructed using motion magnification from facial areas where skin is visible. Due to its design and construction, the FBF does not need information from the eyes, nose, or mouth, and, therefore, it yields a versatile biometric of great potential. In this work, we evaluate the effectiveness of novel Temporal Partitioning and FFT-based features that capture the temporal evolution of facial blood flow. These new features, along with a "time-distributed" CNN-based deep learning architecture, are experimentally shown to increase the performance of FBF-based person identification, compared to our previous efforts. This study provides further evidence of the FBF's potential for use in biometric identification.

Keywords: biometrics; motion magnification; facial blood flow

Citation: Raia, M.; Stogiannopoulos, Th.; Mitianoudis, N.; Boulgouris N.V. Person Identification using Temporal Analysis of Facial Blood Flow. *Electronics* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

1. Introduction

In recent decades, the importance of biometric identification has been increasing steadily. The biometric characteristics of people and their use in recognition systems are now popular for the authentication of individuals [1]. Most mobile devices employ traditional biometric traits, including fingerprints and face, for the verification of the identity of their user. The face and the visual facial characteristics have always been used for identity verification with great success [2–5]. A constraint in common face recognition systems is that they may be sensitive to facial expression variations [4], especially when faces are captured in the wild. Fingerprints have also been widely used in person identification in current mobile phones and police forensics [6–8], but their operation relies on the proximity of the sensing probe to the acquired fingerprint.

The research community always seeks novel biometric technologies in order to complement traditional biometric systems and increase their accuracy [9]. Apart from traditional biometric traits, which have been thoroughly tested and deployed, new biometric modalities have also been devised and explored, such as palm [10], vein [11], ear [12], eye blinking [13] or EEG [14]. These biometrics are either deployed in stand-alone identification applications, or are used as part of multi-modal biometric systems [1]. In such multi-modal cases, the novel biometric traits are usually complementary to traditional biometrics, aiming to enhance end-to-end system performance.

Video amplification was first introduced by Wu et al. [15] and aimed at amplifying, and making visible to the human eye, small motions in videos captured using an ordinary camera. In [15,16], it was demonstrated that, by applying video amplification on facial image sequences, it is possible to visualise the flow of facial blood. In [17], a complex steerable pyramid decomposition was employed and motion amplification was applied solely on the phase component of the decomposition. This enabled motion amplification to focus on the edge information that is prevalent in the phase component. Due to its ability to reveal imperceptible information, motion magnification has been applied to various fields, ranging from computer vision and biomedical imaging to civil and mechanical engineering [18–21].

The widespread use of masks in public places during the COVID19 pandemic opened another possible application domain for motion magnification. Facial masks cover most facial areas that are normally used by efficient face recognition systems. In a situation where part of the face is covered, it would be most useful to devise a facial biometric trait that focuses on the less textured facial areas and does not rely on the visibility of conventional facial landmarks, such as eyes, nose, mouth and eyebrows. In addition, any new facial biometric should be robust to facial expression variations. For the above reasons, we will not be comparing traditional face biometrics with the proposed facial blood flow biometric trait.

Facial blood flow has seen a lot of medical applications recently [22–24]. Inspired by the application of video magnification to the visualization of subtle facial motion, we embarked on exploring the use of facial blood flow patterns as a biometric trait. To that end, we sought evidence that suggests that facial blood flow can act as a biometric. In [25], Buddharaju et al. explored the possibility of performing face identification using thermal infrared cameras. In their study, they support, based on medical evidence and their conducted experiments, that the facial network of veins and arteries can have great variability among individuals, and thus it can act as a distinguishing trait. In addition, they show that the extraction of such facial traits can have repeatability in individuals over multiple days and, as such, they can serve as a biometric.

In our previous work, we investigated the use of Facial Blood Flow (FBF) as a potential biometric trait. The method we developed uses commercial RGB cameras (24 bit Color cameras with HD resolution (1920×1080 pixels) and 25-30 fps) and performs image processing and motion magnification to detect and visualize facial blood flow patterns that can be used for biometric identification. In [26], we proposed a baseline method that used Facial Blood Flow (FBF) for person identification. In that method, FBF was extracted from a person's face using a commercial RGB camera. We used motion amplification [15] to enhance and reveal the actual blood flow in common RGB video streams. Our method in [26] was a contactless method that did not utilize any traditional facial features, i.e., eyes, nose, mouth, eyebrows. Instead, it extracted small facial areas that are not commonly obstructed by facial hair, and it used the motion-amplified video to extract spatio-temporal blood flow information. That approach was shown to work well as a distinctive biometric [26]. Unlike the work in [25], our method does not need a high-grade infrared camera to capture the vein-arteries structure of the face. Instead, our approach uses a low-cost commercial RGB camera to capture three facial areas and use image processing to construct the FBF. In [27], we improved our baseline method by taking into account the temporal evolution of FBF within a period, which was not explicitly considered in [26]. The importance of temporal evolution was stressed before in the literature [28,29] Furthermore, we adopted a deep Convolutional Neural Network (CNN) architecture, which improved the accuracy of the baseline system.

In this paper, we propose an improved new framework, based on several additional features that exploit the temporal evolution of FBF, i.e., temporal partitioning and FFT-based features. Further, we examine a number of different deep learning architectures for the classification task. Our ablation study over different features and network architectures yields that the final proposed system achieves efficient person-identification. The proposed system features a 'time-distributed' CNN architecture to capture the temporal evolution of FBF.

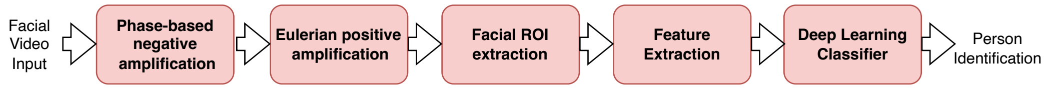


Figure 1. The proposed person identification system based on Facial Blood Flow (FBF).

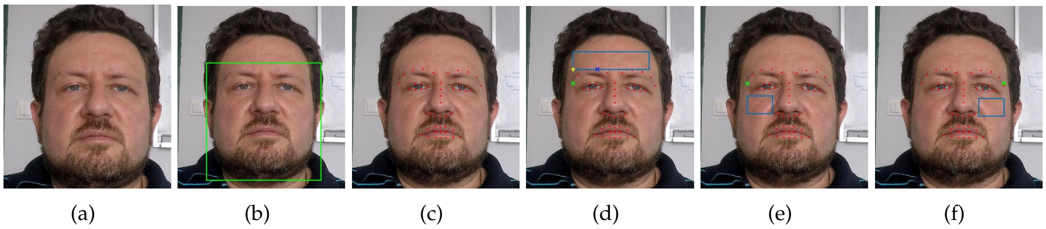


Figure 2. (a) Original face image, (b) Face Detection using [30], (c) Active Appearance Model (AAM) fit using [31]. The three control points (two from the AAM and another inferred from the other two) are highlighted, (d) Detection of the forehead region using two control points, (e), (f) Detection of the left and right facial regions using the left and right control points respectively. (The subject in this figure has agreed to have his image included in the paper for demonstration purposes.)

2. A Facial blood-flow biometric system

This section describes the individual subsystems of the proposed facial blood-flow biometric system. The proposed system is depicted in Fig. 1. In the following subsections, each individual module of the system is described in detail.

2.1. Video capture

The proposed system uses a RGB camera in order to capture the Facial Blood Flow (FBF) of a person. FBF is a periodical phenomenon with period equal to the subject's heart rate. The common human pulse rate at rest varies between 60 to 120 bpm (beats per minute), which translates to frequencies between $1Hz$ and $2Hz$. Thus, a commercial RGB camera with 30 fps (i.e., $30Hz$) video capture will be sufficient to capture several frames from each cycle of facial blood flow. Although video resolution is less important, modern commercial cameras offer HD quality video capture, which is most suitable for our target application. More details regarding the data collection process are provided in the experimental results section.

2.2. Facial Blood Flow Motion Amplification

The next step in the system is to perform motion amplification in the input video. Motion amplification is a remarkable algorithm that amplifies subtle movements in an RGB video. The algorithm detects and amplifies motion, and adds amplified versions of the motion to the original video in order to enhance it visually [15]. Motion amplification is the means by which Facial Blood Flow (FBF) is calculated. Therefore, motion amplification is extremely important in our system. In other biometrics (conventional face recognition, gait recognition, or iris recognition), due to the fact that their respective features are directly visible, motion amplification is not required.

In [15], Wu et al. proposed an Eulerian motion amplification approach, where each image frame from the input video is decomposed into a multi-scale Laplacian pyramid and motion amplification is applied to each scale before reconstruction. In order to amplify movements that exist around object boundaries, i.e., image edges, Wadhwa et al. [17] proposed a phase-based motion amplification scheme, where each frame from the input video is decomposed into a complex steerable pyramid, where the amplification is applied to the phase component of the decomposition only. This approach exploits the well-known fact that edge information in images is best represented by the phase information in the Fourier domain [32]. This approach achieves superior motion amplification around object edges and boundaries, as opposed to the Eulerian motion amplification which is more efficient in enhancing motion that exists within the texture of an object.

The motion amplification algorithm will not only amplify subtle motion taking place within the object, but will also amplify the subject's global motion. Hence, if the subject is moving as a whole, its global movement will also be amplified, producing output that is not useful for our application. Therefore, an important requirement for the efficient function of motion amplification is that the captured subject must be as still as possible, exhibiting minimal global movement. For that reason, in our application we capture human subjects who have been advised to stay as motionless as they can. In practice, however, it is unavoidable that there will be some slight global movement during capturing. Due to that, if motion amplification was used directly on the original video, the output would not be particularly informative regarding subtle facial movements. To avoid such problems, we apply pre-processing in order to filter out unavoidable small global motions before motion amplification is applied.

As a first processing step, we use the phase-based motion amplification [17] with a negative amplification factor α, in the range of $[-1, 0)$, in order to reverse any unwanted global motion. The actual value of α was determined experimentally. As explained earlier, the phase-based motion amplification is more sensitive to edge movements that are due to unwanted small global movements, observed when the subject is moving as a whole. Therefore, applying negative amplification using the phase-based motion amplification algorithm will stabilise the subject within video as much as possible without affecting the subtle movements in the facial texture area, which is where the facial blood flow can be observed.

Subsequently, we use the Eulerian Video Magnification method by Wu et al. [15] to amplify the facial blood-flow. The Eulerian video magnification method tends to magnify the image value oscillations, thus it has proven [15] to be ideal for amplifying blood circulation. In contrast to [15], where a new (motion-enhanced) video sequence was created, i.e., the amplified motion is added to the original video, in order to create a highlighted video, our focus is only on the extraction of the amplified facial blood flow. Consequently, we extract the amplified motion, without adding it to the original video, as in [15]. In order to keep computational complexity low, we apply motion amplification to a grayscale version of the input video. The result of this procedure is the calculated facial blood flow of a subject, represented in the form of an image sequence $\mathcal{Q}(x, y, t)$, where x, y denote the image coordinates and t the frame index. Here, we should stress that without motion magnification, the signal $\mathcal{Q}(x, y, t)$ is zero, therefore there are no extracted features to model. Finally, in our approach we only use grayscale data. The reason is that different colour planes would produce different apparent motions, which would lead to inaccurate facial blood flow calculation. Therefore, using aggregate information in the form of grayscale pixel intensities is a safer option.

2.3. *Extraction of facial regions of interest*

To verify the validity of FBF as a biometric, we rely on facial areas that do not contain facial landmarks, which are used in conventional face biometry. Traditional facial recognition biometric approaches use the entire face, including the eyes, nose and mouth, the shape of which are specific to each individual and, therefore, have significant discriminatory capacity. Instead, in our approach, we consider the use of facial areas that do not include discriminatory landmarks, but are certain to involve substantial blood flow that can be revealed. With that in mind, we focused on three selected areas of interest: a) the lower forehead area, b) the skin area below the left eye, and c) the skin area below the right eye. These areas feature mainly facial skin without landmarks, and are ideal for FBF acquisition. This selection of facial areas is supported by the findings in the work of Buddharaju et al. [25], which is an exploration of the facial areas containing veins and arteries that can be used for person identification. As it can be seen, the forehead and the two left and right cheek areas contain multiple veins and arteries, a fact that makes them particularly suitable for the observation and representation of facial blood flow. This finding is also supported in the work by von Arx et al. [33], where a complete atlas of the facial blood vessels is presented. Another advantage of our chosen facial areas is that they are always visible or can easily become visible and easy to capture. The special situation where only the forehead area is visible is examined as a separate experiment. Such a situation arises when faces are partly covered by a mask.

To detect these regions, we use the Zhu-Ramanan face detector [30] to isolate the facial area. Then, we fit an Active Appearance Model (AAM) [34] with 68 facial features using the implementation of Bulat and Tzimiropoulos [31]. Using the facial features, inferred by the AAM, we select an area of 71×201 pixels above the eyes and eyebrows to extract the forehead area. The left and right facial areas are determined by positioning two 71×101 rectangles below the eyes.

The exact positions of the left and right facial areas are calculated by exploiting the position of key-points on the eyes and eyebrows. In Fig. 2(d) we depict the eyebrows' highest and rightmost points using blue and green points respectively. Using this information we can infer the point where the forehead commences above the eyebrows. The x-coordinate of the eyebrows' highest point and the y-coordinate of the eyebrows' rightmost point are used to form the starting point of the forehead, which is shown as a yellow cross in Fig. 2(d). The width and height of the forehead rectangle are fixed, thus extending the rectangle over both eyebrows. In a similar fashion, we exploit the leftmost eyebrows' point as a reference point (green point in Fig. 2(e)), and proceed down by 120 pixels so as to detect the bottom left corner of the rectangle for the left facial. Equally, using the rightmost eyebrow point, we can set a rectangle containing the right facial area (see Fig. 2(f)). Fig. 2 shows an example of the rectangle positioning procedure. The sizes of the rectangles were set based on experimentation. An alternative approach would be to deploy image registration algorithms to fully automate rectangle positioning. Nonetheless, in order to keep the setup as simple as possible, and not introduce distortions due to image transformation, we kept these sizes constant and asked the participants to sit at fixed distance from the camera. This ensured that minimal registration problems would appear. The AAM for the facial structure can provide sufficient information to tackle the face registration problems that would arise if the algorithm is applied in less controlled environments [35]. Snapshots from the three extracted videos from the three areas are depicted in Fig. 3.

(a) Forehead

(b) Left area

(c) Right area

Figure 3. Snapshots from the three extracted areas for the subject in Fig. 2. It is clear that these areas don't contain any traditional facial biometric traits.

2.4. Temporal and Spectral Feature Extraction

The next step is to extract efficient feature structures that will enable robust person identification. In order to achieve faster and more practical identification, we limit the duration of the clips that will be presented to the deep learning classifier to $T = 1$ sec, so as to capture at least one cycle of blood circulation. Nonetheless, as it is not always easy to discern the start, end, or other stages within the facial blood cycle, synchronization issues may occur. This highlights the importance of extracting features that are robust to phase variations within the observed periodic phenomenon. In [26,27], we proposed two baseline methodologies, which we extend here and show that they can lead to increased performance.

2.4.1. Temporal features

In [26], we proposed to average all frames in each video clip, creating the average frame that was used for training and recognition. Although that approach provided robustness to phase variations, it ignored the temporal evolution of the FBF within the clip. Thus, in this paper, we propose a temporal partitioning, i.e., divide each clip into five sub-clips of equal duration T_s, and calculate the average FBF for each sub-clip separately. Assuming that we have a video sequence $\mathcal{Q}(x,y,t)$, where x,y are the spatial co-ordinates and t represents the temporal evolution, the proposed features $F_{temp}(x,y,k)$ can be described, as follows (see Fig. 4):

$$F_{temp}(x,y,k) = \frac{1}{T_s} \sum_{t=(k-1)T_s}^{kT_s} \mathcal{Q}(x,y,t), \quad 1 \leq k \leq 5 \tag{1}$$

The average image and the temporal averages for the FBF biometric, extracted from the forehead region of two subjects, are depicted in Fig. 5 and 6 respectively. These templates clearly show the different FBF patterns in different individuals and, therefore, they demonstrate the discriminatory capacity of the FBF biometric.

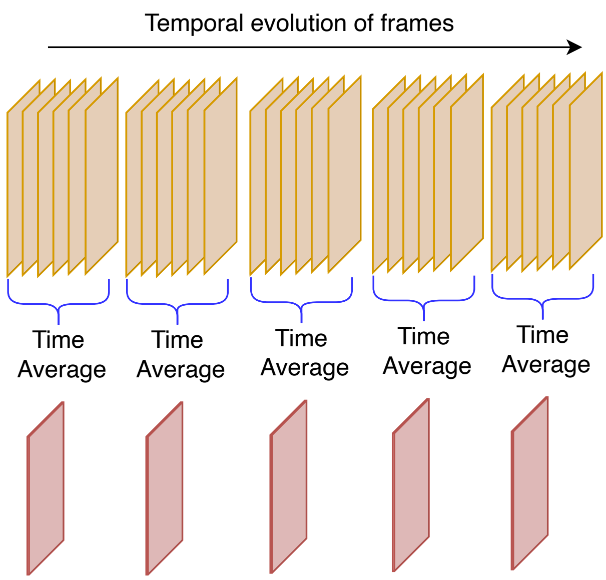


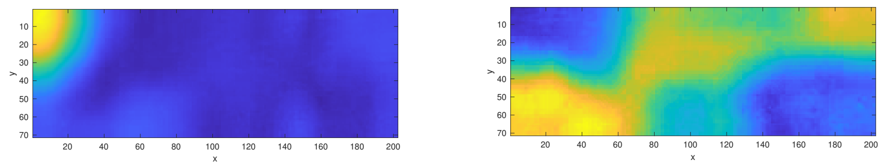
Figure 4. Division of a video clip into five sub-clips. Temporal averages are calculated for each sub-clip.

2.4.2. Transform-domain phase-invariant features

In [27], we proposed to use the 1D-Discrete-Cosine Transform (DCT) along the temporal axis of the sequence $\mathcal{Q}(x,y,t)$, i.e.

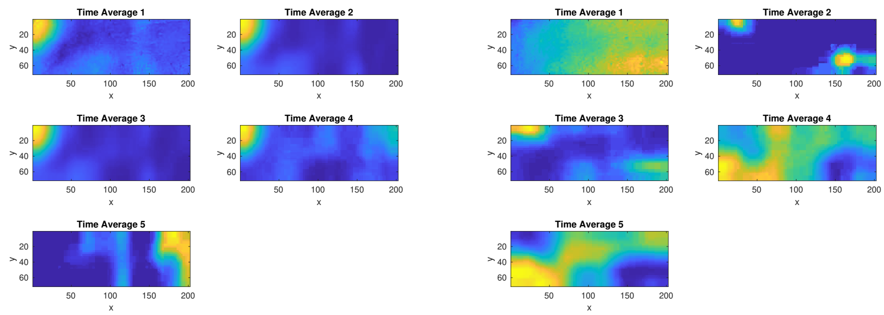
$$F_{DCT}(x,y,k) = \sum_{t=0}^{T} \mathcal{Q}(x,y,t) \cos\left[\frac{\pi k}{T}(t+\frac{1}{2})\right], 0 \leq k \leq T \tag{2}$$

The above DCT features provided a viable solution, since the DCT is real-valued and approximately phase-invariant.



(a) Subject A (b) Subject B

Figure 5. Temporal features for the FBF biometric extracted from the forehead region. The averaged image template [26] is shown for for subject A and B. Lighter colors represent greater values while darker colors represent smaller values (best seen in color).



(a) Subject A (b) Subject B

Figure 6. Temporal features for the FBF biometric extracted from the forehead region. The proposed temporal averages of (1) is shown for subject A and B. Lighter colors represent greater values while darker colors represent smaller values (best seen in color).

In the present work, however, we propose to use the abs-magnitude of the Fast-Fourier Transform (FFT), since the FFT is always phase-invariant. In addition, because of the magnitude FFT's even symmetry, the negative frequency components of the transformation can be dropped, yielding smaller-sized features. The proposed feature is given by:

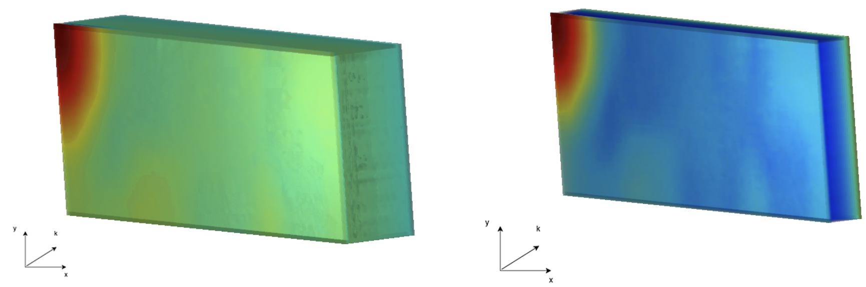
$$F_{FFT}(x,y,k) = \left| \sum_{t=0}^{T} \mathcal{Q}(x,y,t) e^{-j\frac{2\pi tk}{T}} \right|, \qquad 0 \leq k \leq T \tag{3}$$

As in the case of DCT features, the proposed FFT features are calculated by applying the FFT transform along the temporal dimension of the input sequence. The two transform-domain features are depicted in Fig. 7 for the forehead region. In either case, the extracted features are 3D matrices with the first two dimensions representing the image spatial coordinates (pixel location), and the third dimension representing the transform-domain content of each pixel (pixel value).

2.5. Deep Learning Architectures for classification

The next step is to use a deep learning architecture in order to perform person identification. As detailed in the previous subsection, the extracted feature data are in the form of 3D cubes which represent spatio-temporal features (from temporal partitioning) or spatio-frequency features (DCT or FFT features for each pixel location). In either case, the input is a 3D tensor of size $M \times N \times K$, where $M \times N$ is the image resolution and K is the number of possible values for each feature.

Many deep learning architectures were investigated to model the evolution in time or frequency of the input features, including 3D CNN [36] and Long-Short Time Memory (LSTM) modules [37]. Nonetheless, in our experiments the most successful was an architecture that included 'time'-distributed 2D CNN modules [38]. In other words, it is



(a) DCT features [27] (b) FFT features

Figure 7. Frequency-domain features for the FBF biometric extracted from the forehead region of subject A. (a) DCT features [27], (b) FFT-features calculated using (3).

a CNN structure, consisting of 2D filters that are applied on the spatial co-ordinates x, y and remain unchanged over the z-axis, which describes time or frequency, depending on whether the spatio-temporal or spatio-frequency 3D feature is used. The proposed architecture is depicted in Figure 8. As seen, it consists of a smaller VGG16-like [39] structure of four 'time'-distributed 2D convolutional layers, followed by two fully-connected layers for classification. More specifically, the first 'time'-distributed 2D CNN contains 16 2×2 filters, the second 32 2×2 filters, which are then followed by a 2×2 max pooling layer. Consequently, two more CNN layers of 32 2×2 filters each and a 2×2 max pooling layer. ReLUs are used after each CNN layer and the Dropout regularisation method [40] using a parameter value of 0.2 is used after each max pooling layer. The features are flattened and presented to two fully-connected layers (FCN or Dense), the first with 64 nodes and the second, being the output layer, with 13 nodes, equal to the number of people to the dataset. After the first FCN, Batch Normalisation (BN) [41] is used along with ReLU and the Dropout regularisation method with a parameter value of 0.3. The output of the last layer is presented to the Softmax activation function. The size of the input is not fixed in order to accommodate the features from the three areas of interest, which are of different size. The parameter values for dropout, and the number and size of filters at each layer, were determined based on experimentation.

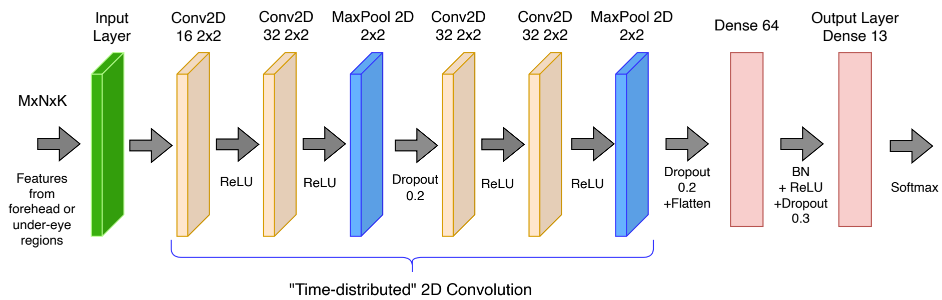
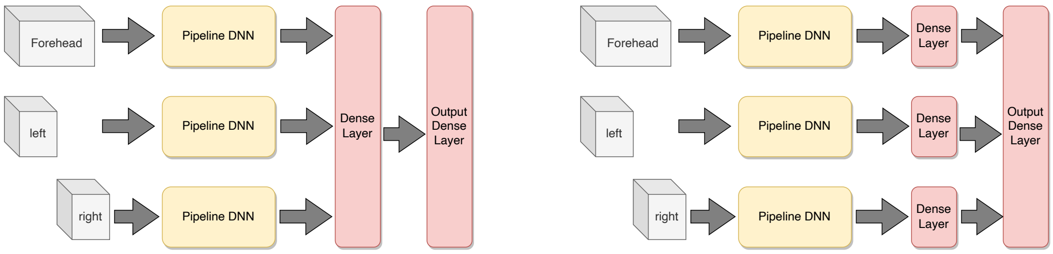


Figure 8. The proposed CNN structure with 'time'-distributed 2D convolutions used in the convolutional layers of the network. Conv2D refers to a 'time'-distributed 2D convolutional layer, ReLU and Softmax refer to the corresponding activation functions, Dropout refers to Dropout regularisation [40], BN refers to Batch Normalisation. The number of filters and the size of the filters is indicated at the top of the respective level.

The system should also combine features from three different facial areas in order to perform person identification. In [27], we determined that the features should be processed separately by the convolutive architecture, proposed in Fig. 8, without the Dense layers (pipeline) and the final features should be concatenated and be presented to a single fully-connected (Dense) stage for classification (Ensemble 1) (see Fig. 9(a)). The Dense Layer consists of 64 weights. Fig. 9(b) depicts a second method (Ensemble 2), where the features from each facial area are independently processed by the proposed convolutive architecture and a separate Dense stage, before being concatenated to the final FCN stage. Each of these Dense stages has 64 nodes. The output dense layer has 13 nodes, equal to the number of subjects in the dataset.



(a) Ensemble 1 (b) Ensemble 2

Figure 9. Two ensemble methods to combine the features from the three facial regions of interest. Pipeline refers to the architecture of Fig. 8 without the fully-connected (Dense) layers.

Table 1. Identification Accuracy for the validation dataset for the three different features and various scenarios. Salt-n-pepper augmentation on FFT features and the combination of the three areas with Ensemble 2 yields the best performance. Values in bold denote the best performance.

Features	Forehead only	All areas		
		Ensemble 1	Ensemble 2	
			No augment.	With augment.
Average Image [26]	68.76%	77.53%	79.72%	79.79%
DCT Features [27]	73.19%	82.61%	84.66%	84.78%
Temporal Partitioning	72.54%	84.51%	85.56%	85.6%
FFT Features	74.03%	85.02%	87.20%	**89.04%**

3. Experiments

3.1. Dataset - Implementation

To test the efficiency of the proposed Facial Blood Flow (FBF) biometric, we created a new dataset. A GoPro Hero 4 Black camera was used for video recording at 1920×1080 resolution and 30 frames per second. Facial image sequences from a total of 13 subjects were recorded in a room with natural light. This was needed because the motion magnification algorithm tends to magnify subtle, nearly invisible, movements and oscillations produced by the flickering of artificial light, which would result in contaminated output. Subjects were asked to sit on a chair at a fixed distance ($\sim$ 1m) from the camera and to remain as motionless as possible during the recording. This guidance was meant to prevent the occurrence of registration problems due to incidental body movement. The camera was positioned at the same height for all subjects, almost at eyes level, in order to minimise registration issues. Eighteen (18) recordings were captured from each subject, each lasting 20 seconds. These recordings were captured for each subject over two different days. This produced a total of 360 seconds of recording per subject, which was used for training. A week later, three additional recordings per subject were captured on a single day. The additional recordings amounted to a total of 60 seconds per subject, and were used for system testing. Conducting the recordings over three different days strengthens our confidence in the temporal robustness of the proposed biometric trait and in its appropriateness for practical use. All recordings were segmented into 1-second clips, yielding 360 training clips and 60 validation clips for each subject. The compiled data can be made available upon request. An example of the captured video frames can be seen in Fig. 2. Video capturing has been performed in a controlled environment, ensuring that no artificial light was used and that the subjects were instructed to remain as still as possible. Although this may seem as a limitation of the method, it is a realistic constraint to have a controlled environment for identity verification, e.g. in a police station or at border control.

The Movement Motion Attenuation and the Facial Blood Flow Amplification algorithms were implemented using MATLAB code, provided by [17] and [15] respectively. For the motion attenuation module, the value of $\alpha = -0.75$ was used, as it was experimentally seen to yield satisfactory results. The amplification factor was set to $\alpha = 120$, while the frequency range of amplification was set between $0.83 - 1$ Hz, as suggested in [15]. The face isolation and AAM fitting stages were based on the method in [31]. The proposed deep architectures were developed in Python using TensorFlow on a PC with an NVidia RTX A6000 GPU, running Ubuntu Linux 20.04[1]. Stochastic Gradient Descent (SGD) was used to train the deep network, using categorical cross entropy as a cost function, whereas the learning rate was set to 0.01 and momentum was set to 0.9. The network was trained for 30 epochs using a batch size of 64 samples.

3.2. Results - Ablation Study

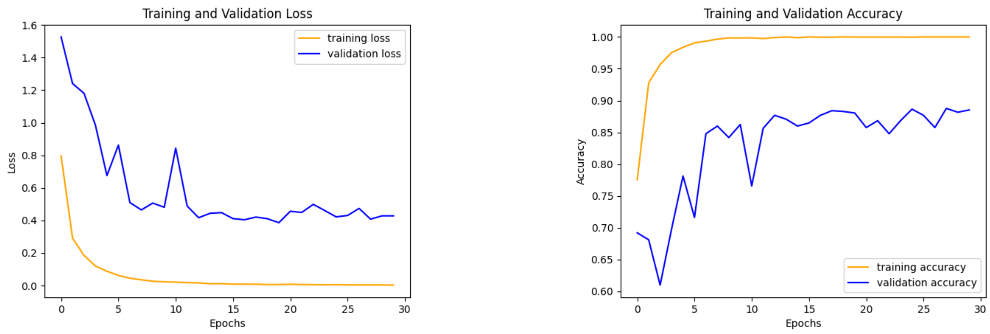
As it was shown in [26], using only forehead features for person identification yields inferior performance in comparison to using the combination of all three proposed facial areas. In the present work, we re-evaluate the use of the forehead area in conjunction with our new FBF representations and the architecture shown in Fig. 8. Results are shown in Table 1. As seen, similar to [26], the forehead area alone still exhibits moderate biometric performance.

Subsequently, we evaluated the combination of the three facial areas, using the two combination approaches, shown in Fig 9. In Table 1, we present the identification accuracy achieved by the two combination methods. Clearly, the Ensemble 2 architecture yields the best results regardless of the feature used. Therefore, it seems that the most efficient strategy is to process each feature through individual FCNs before combining features in the output (final) layer. In addition, it is clear from Table 1 that the inclusion of the temporal element surely improves performance. All architectures that incorporate the temporal element (DCT Features, Temporal Partitioning and FFT Features) perform better than using the Average Image.

Finally, we experimented with data augmentation. We attempted three types of augmentation that seemed suitable for our problem: Gaussian noise, Salt-n-pepper noise, and Vertical flipping. Noise addition was considered because it was shown to improve network generalisation and reduce overfitting [42]. Out of the three options, only the addition of salt-n-pepper noise to the final extracted features appeared to improve performance. Specifically, the saturation of 10% of each image pixel to equal percentage of black and white pixels seemed to yield performance improvement. This augmentation doubled the available training data. Results are shown in Table 1 in terms of Classification Accuracy.

As seen in Table 1, augmentation with salt-n-pepper noise improved the performance of all feature sets. Further, the proposed FFT features are consistently superior in all scenarios, reaching a recognition rate equal to 89.04%. The Temporal Partitioning features rank second in performance, followed by the DCT Features [27] and the Average Image feature of [26]. The performance achieved by our new features and architecture is an improvement over the system in [27], and shows the potential of the FBF biometric trait for effective person identification. The evolution of the loss function and accuracy over 30 epochs for the proposed "time-distributed" VGG network with the FFT-features and the Ensemble 2 strategy is depicted in Fig. 10. The confusion matrix for the proposed architecture and features is shown in Fig. 11. It is clear that the system does not overfit and in addition the classification performance is robust since only minor mis-classifications are observed.

[1] The code is available at https://github.com/mitia98/FBF_person_id



(a) Loss function (b) Accuracy

Figure 10. The evolution of the loss function and accuracy over 30 epochs for the proposed "time-distributed" VGG network with the FFT-features and the Ensemble 2 strategy.

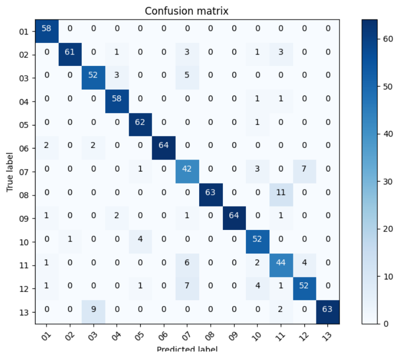


Figure 11. The confusion matrix for the proposed "time-distributed" VGG network with the FFT-features and the Ensemble 2 strategy.

In Table 2, we present an ablation study with other deep architectures that were examined in our investigation. The study was performed using the FFT features from all regions of interest and the Ensemble 1 architecture in terms of Accuracy and False Positive Rate. The study included 2D-CNN VGG, 3D-CNN VGG, the "time-distributed" VGG of Fig. 8, stacked LSTM and the "time-distributed" VGG with LSTM layers. The exact details of these architectures are not presented here, due to limited interest, but they are of similar size to the one presented in Fig. 8, and therefore are a good basis for direct comparisons. From the results presented in Table 2, it is clear that the inclusion of LSTM modules did not improve performance. Instead, a more shallow "time-distributed" VGG is more efficient.

Table 3 presents comparisons in terms of computational complexity. Complexity is measured in terms of the required number of network parameters, the network size (in MBs), and the average inference time per batch. The compared systems are the afore-mentioned reference architectures, which use FFT features from all regions of interest, and the Ensemble 1 architecture. In terms of parameters and size, the Time-Distributed CNN (VGG version), which achieved the highest identification performance, is the largest network, with 55.7 million parameters requiring 435.6 MB of storage. In terms of inference time, that architecture ranks in the middle. Although inference times were measured in a high-specification machine, they are very small and can comfortably support a practical real-time identification scenario. In particular, the per-sample inference time achieved by the best-performing Time-Distributed CNN network is equal to $0.742/64 = 11.6$ ms, where the total inference time was divided by the number of samples (64) in the batch. Such low inference times make our proposed application viable for use even on everyday computational platforms (e.g., common PCs).

Table 2. Comparison between different Ensemble-1 architectures, in conjunction with FFT features extracted from all facial regions of interest. No data augmentation was applied. Values in bold denote the best performance. The proposed "time-distributed" VGG architecture yields the best performance.

Model	Accuracy	False Positive Rate
2D-CNN VGG version	71.01%	2.6%
3D-CNN VGG version	81.26%	1.64%
Time Distr. CNN VGG version	**85.02%**	**1.27%**
Stacked LSTM	78.05%	1.62%
Time Distr. CNN VGG version + LSTM	79.59%	1.51%

Table 3. Comparison in terms of computational complexity, presenting the number of parameters, model size, and inference time (per batch) for each of the examined architectures. The proposed "time-distributed" VGG architecture has the greatest number of parameters and model size, nonetheless it doesn't require the greatest inference time per batch.

Model	No. of Parameters	Model Size (MB)	Inference Time per batch (sec)
2D-CNN VGG version	7M	54.7	0.733
3D-CNN VGG version	13.9M	109.3	0.736
Time Distr. CNN VGG version	55.7M	435.6	0.742
Stacked LSTM	14.89M	116.4	0.771
Time Distr. CNN VGG version +LSTM	14.03M	109.9	0.786

4. Conclusions

In this paper, we presented evidence of the validity of Facial Blood Flow (FBF) as a biometric trait. We examined two new features: the FFT features and the temporal partitioning features, which yield improved performance over the performance achieved using the DCT features we proposed in our past work. In the present paper, We also proposed a new 'time'-distributed CNN architecture that consolidates the performance gains using the new features. The presented algorithms and experiments demonstrate the effectiveness of FBF as a stand-alone or a complementary biometric trait and serve as a well-founded proof-of-concept. In our future work, we will expand our experiments using larger datasets and multimodal features. More specifically, we are planing to study the performance of the FBF biometric on a more diverse database, exploring also the use of multi-spectral cameras. In future research, 3D modelling can be used for very accurate compensation of incidental head movements before blood flow calculation is performed.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, N.M. and N.B.; methodology, N.M. and N.B.; software, M.R.; validation, M.R., T.S. and N.M.; formal analysis, N.M. and N.B.; investigation, M.R.; writing—original draft preparation, M.R. and T.S.; writing—review and editing, N.M. and N.B.; visualization, M.R. and T.S.; supervision, N.M. and N.B.; project administration, N.M. All authors have read and agreed to the published version of the manuscript.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: "This research received no external funding"

Institutional Review Board Statement: "Ethical review and approval were waived for this study due to the anonymization of the final data used by the proposed algorithm."

Informed Consent Statement: Written informed consent has been obtained from all participants in the study, including the patient(s) whose information is included in this paper. The consent form includes a full explanation of the nature and purpose of the study, as well as any potential risks and benefits of participation. The participants were informed that their participation is voluntary, and that they have the right to withdraw from the study at any time without any negative consequences. The participants were also informed that their information will be kept confidential and anonymous, and that the data collected will only be used for the purposes of the research study

Data Availability Statement: We encourage all authors of articles published in MDPI journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required. Suggested Data Availability Statements are available in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics.

Acknowledgments: We gratefully acknowledge the support of NVIDIA Corporation with the donation of the RTX A6000 GPU used for this research. We would also like to thank the 13 students at the Electrical and Computer Engineering Department at Democritus University of Thrace, who participated in our experiment.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ross, A.A.; Nandakumar, K.; Jain, A.K. *Handbook of Multibiometrics*; Springer: New York, NY, 2006; p. 220.
2. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, Present, and Future of Face Recognition: A Review. *Electronics* **2020**.
3. Taskiran, M.; Kahraman, N.; Erdem, C.E. Face recognition: Past, present and future (a review). *Digital Signal Processing* **2020**, *106*, 102809.
4. Oloyede, M.O.; Hancke, G.P.; Myburgh, H.C. A review on face recognition systems: recent approaches and challenges. *Multimedia Tools and Applications* **2020**, *79*, 27891 – 27922.
5. Li, L.; Mu, X.; Li, S.; Peng, H. A Review of Face Recognition Technology. *IEEE Access* **2020**, *8*, 139110–139120.
6. Yang, W.; Wang, S.; Hu, J.; Zheng, G.; Valli, C. Security and Accuracy of Fingerprint-Based Biometrics: A Review. *Symmetry* **2019**, *11*, 141.
7. Yao, Z.; Bars, J.M.L.; Charrier, C.; Rosenberger, C. Literature review of fingerprint quality assessment and its evaluation. *IET Biom.* **2016**, *5*, 243–251.
8. Kaushal, N.; Kaushal, P. Human Identification and Fingerprints: A Review. *Journal of biometrics & biostatistics* **2011**, *2*, 1–5.
9. Ghayoumi, M. A review of multimodal biometric systems: Fusion methods and their applications. *2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS)* **2015**, pp. 131–136.
10. Liang, X.; Yang, J.; Lu, G.; Zhang, D. CompNet: Competitive Neural Network for Palmprint Recognition Using Learnable Gabor Kernels. *IEEE Signal Processing Letters* **2021**, *28*, 1739–1743.
11. Kuzu, R.S.; Maiorana, E.; Campisi, P. Vein-Based Biometric Verification Using Densely-Connected Convolutional Autoencoder. *IEEE Signal Processing Letters* **2020**, *27*, 1869–1873.
12. Nikose, S.; Meena, H.K. Ear-biometrics for human identification. *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)* **2020**, pp. 8–13.
13. Abo-Zahhad, M.; Ahmed, S.M.; Abbas, S.N. A Novel Biometric Approach for Human Identification and Verification Using Eye Blinking Signal. *IEEE Signal Processing Letters* **2015**, *22*, 876–880.
14. Alsumari, W.; Hussain, M.; AlShehri, L.; Aboalsamh, H. EEG-Based Person Identification and Authentication Using Deep Convolutional Neural Network. *Axioms* **2023**, *12*, 74.
15. Wu, H.Y.; Rubinstein, M.; Shih, E.; Guttag, J.; Durand, F.; Freeman, W.T. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Trans. on Graphics* **2012**, *31*.
16. Rubinstein, M. Analysis and Visualization of Temporal Variations in Video. PhD Thesis, Massachusetts Institute of Technology, 2013.
17. Wadhwa, N.; Rubinstein, M.; Durand, F.; Freeman, W. Phase-based Video Motion Processing. *ACM Trans. on Graphics* **2013**, *32*.
18. Lin, Z.; Chen, X.; Chen, G. Cervical Pulse Wave Extraction Based On Video Motion Amplification. *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)* **2021**, pp. 784–787.
19. Śmieja, M.; Mamala, J.; Prażnowski, K.; Ciepliński, T.; Szumilas, L. Motion Magnification of Vibration Image in Estimation of Technical Object Condition-Review. *Sensors (Basel, Switzerland)* **2021**, *21*.
20. Fioriti, V.; Roselli, I.; Tatì, A.; Romano, S.; Canio, G.D. Motion Magnification Analysis for structural monitoring of ancient constructions. *Measurement* **2018**.
21. Zhang, D.; Zhu, A.; Hou, W.; Liu, L.; Wang, Y. Vision-Based Structural Modal Identification Using Hybrid Motion Magnification. *Sensors (Basel, Switzerland)* **2022**, *22*.
22. Zhang, Y.; Chen, S.; Ruan, Y.; Lin, J.; Li, C.; Li, C.; Xu, S.; Yan, Z.; Liu, X.; Miao, P.; et al. The Facial Skin Blood Flow Change of Stroke Patients with Facial Paralysis after Peripheral Magnetic Stimulation: A Pilot Study. *Brain Sciences* **2022**, *12*.

23. Yoshida, K.; Nishidate, I. Phase Velocity of Facial Blood Volume Oscillation at a Frequency of 0.1 Hz. *Frontiers in Physiology* **2021**, *12*.
24. Fu, G.; Zhou, X.; Wu, S.J.; Nikoo, H.M.; Panesar, D.; Zheng, P.P.; Oatley, K.; Lee, K. Discrete emotions discovered by contactless measurement of facial blood flows. *Cognition and Emotion* **2022**, *36*, 1429 – 1439.
25. Buddharaju, Pradeepand Pavlidis, I.M.C., Face Recognition Beyond the Visible Spectrum. In *Advances in Biometrics: Sensors, Algorithms and Systems*; Ratha, Nalini K.and Govindaraju, V., Ed.; Springer London: London, 2008; pp. 157–180.
26. Pistola, T.; Papadopoulos, A.; Mitianoudis, N.; Boulgouris, N.V. Biometric Identification using Facial Motion Amplification. In Proceedings of the Proc. of IEEE Int. Conf. on Image Processing (ICIP), Taipei, Taiwan, 2019.
27. Gkentsidis, K.; Pistola, T.; Mitianoudis, N.; Boulgouris, N.V. Deep Person Identification using Spatiotemporal Facial Motion Amplification. In Proceedings of the Proc. of IEEE Int. Conf. on Image Processing (ICIP), Abu Dhabi, UAE, 2020.
28. Roccetti, M.; Gerla, M.; Palazzi, C.E.; Ferretti, S.; Pau, G. First Responders' Crystal Ball: How to Scry the Emergency from a Remote Vehicle. *2007 IEEE International Performance, Computing, and Communications Conference* **2007**, pp. 556–561.
29. Sun, Z.; Torrie, S.A.; Sumsion, A.W.; Lee, D.J. Self-Supervised Facial Motion Representation Learning via Contrastive Subclips. *Electronics* **2023**.
30. Zhu, X.; Ramanan, D. Face detection, pose estimation and landmark localization in the wild. In Proceedings of the Proc. of Computer Vision and Pattern Recognition (CVPR), Providence, Rhode Island, USA, 2012.
31. Bulat, A.; Tzimiropoulos, G. Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources. In Proceedings of the Proc. of Int. Conf. on Computer Vision (ICCV), Venice, Italy, 2017.
32. Gonzalez, R.C.; Woods, R.E. *Digital image processing*; Prentice Hall: Upper Saddle River, N.J., 2008.
33. von Arx, T.; Tamura, K.; Oba, Y.; Lozanoff, S. The Face – A Vascular Perspective. 2018, Vol. 128, pp. 382–392.
34. Cootes, T.; Edwards, G.; Taylor, C. Active Appearance Models. In Proceedings of the Proc. of European Conference on Computer Vision (ECCV), Freiburg, Germany, 1998.
35. Liu, X. Generic Face Alignment using Boosted Appearance Model. *2007 IEEE Conference on Computer Vision and Pattern Recognition* **2007**, pp. 1–8.
36. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning spatiotemporal features with 3D convolutional networks. In Proceedings of the Proc. of IEEE Int. Conf. on Computer Vision (ICCV), Santiago, Chile, 2015.
37. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735–1780.
38. Koeppe, A.; Bamer, F.; Markert, B. An intelligent nonlinear meta element for elastoplastic continua: deep learning using a new Time-distributed Residual U-Net architecture. *Computer Methods in Applied Mechanics and Engineering* **2020**, *366*.
39. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* **2014**, *abs/1409.1556*.
40. Srivastava, N.; Hinton, G.E.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
41. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the International Conference on Machine Learning, 2015.
42. Grandvalet, Y.; Canu, S.; Boucheron, S. Noise Injection: Theoretical Prospects. *Neural Computation* **1997**, *9*, 1093–1108.