# Unconstrained Biometric Recognition based on Thermal Hand Images

Ewelina Bartuzi
*Biometrics Laboratory*
*Research and Academic Computer Network (NASK)*
Warsaw, Poland
ewelina.bartuzi@nask.pl

Katarzyna Roszczewska
*Biometrics Laboratory*
*Research and Academic Computer Network (NASK)*
Warsaw, Poland
katarzyna.roszczewska@nask.pl

Andrzej Pacut
*Biometrics Laboratory*
*Research and Academic Computer Network (NASK)*
Warsaw, Poland
andrzej.pacut@nask.pl

Adam Czajka
*Department of Computer Science and Engineering*
*University of Notre Dame*
Notre Dame, IN, USA
aczajka@nd.edu

***Abstract*—This paper proposes a biometric recognition method based on thermal images of inner part of the hand, and a database of 21,000 thermal images of both hands acquired by a specialized thermal camera from 70 subjects. The data for each subject was acquired in three different sessions, with two first sessions organized on the same day, and the third session organized approximately two weeks apart. This allowed to analyze the stability of hand temperature in both short-term and long-term horizons. No hand stabilization or positioning devices were used during acquisition, making this setup closer to real-world, unconstrained applications. This required making our method translation-, rotation- and scale-invariant. Two approaches for feature selection and classification are proposed and compared: feature engineering deploying texture descriptors such as Binarized Statistical Image Features (BSIF) and Gabor wavelets, and feature learning based on convolutional neural networks (CNN) trained in different environmental conditions. For within-session scenario we achieved 0.36% and 0.00% of equal error rate (EER) in the first and the second approach, respectively. Between-session EER stands at 27.98% for the first approach and 17.17% for the second one. These results allow for estimation of a short-term stability of hand thermal information. This paper presents the first known to us database of hand thermal images and the first biometric system based solely on hand thermal maps acquired by thermal sensor in unconstrained scenario.**

## I. Introduction

Palm geometry and vein patterns are common biometric modes related to hand, whereas the use of thermal features for identity recognition has not gained too much attention. This can be related to high prices of thermal cameras, although over the years the cost and size of thermal sensors have been decreasing. Portable infrared sensors that can be connected to smartphones recently offered on the market can make this biometric mode more accessible.

The main advantage of using thermal information in biometrics is that spatial distribution of temperature is difficult to reconstruct with a spoofing artifact. This stands for having liveness detection built into this method by default. Adding unconstraint acquisition, thermal-based hand recognition appears as a method that is hygienic, non-invasive, fast, and independent of ambient light. Biometric recognition based on hand thermal features should not be confused with hand geometry or vein pattern recognition. Thermal properties are closely associated to the pattern of blood vessels, but these are separate approaches to identification problem. However, it is possible that multimodal system, which combines these two characteristics may achieve high reliability.

Certainly, we should be aware of a few important limitations related to physiological and environmental factors that influence the body temperature, such as health conditions, emotions, dynamics of metabolic processes, anatomy, muscle function and subject's activity right before data acquisition. This paper is the first work known to us investigating a reliability of hand thermal features in a short-term period (approx. two weeks) acquired by a professional thermal sensor in unconstrained environment. The database collected for $70$ subjects and encompassing $21,000$ thermal images is made also available to facilitate further research in this topic.

This work is organized around three questions that we answer in the next subsections:

Q1. Is the distribution of hand temperature measured by a thermal sensor in unconstrained setup unique?
Q2. Are the thermal features stable and able to offer a reliable biometric recognition after a period of time?
Q3. Which technique of feature extraction and classification applied to thermal images delivers the highest accuracy?

In section II we review past work related to thermal hand features. Section III presents the acquisition protocol and the database of thermal images. In Section IV we present image pre-processing steps required prior feature extraction and classification presented in Section V. Concluding remarks in Section VII include discussion on advantages, limitations and future work.

## II. Related Work

Use of thermal information in biometrics is typically narrowed to either object detection and segmentation or to supporting technique for non-thermal feature extraction. The latter is especially effective when the background temperature is substantially different from objects, such as faces, hands or veins. There are different ways to measure the temperature used to visualize the heat distribution. The measurement can be carried out with sensors using changes in the resistance of the object under the influence of temperature, changes in the volume of fluids or by the registration of the thermal radiation emitted by a human body, which emits waves from the range of $\langle 2\mu m; 20\mu m \rangle$.

Thermal images of the back of the hand are usually employed for identity recognition based on the veins pattern, thanks to high efficiency of biometric systems based on the hand vein patterns. Attempts to obtain vein patterns from thermal images date back to the mid-1990s [1]. In these solutions, images are usually obtained using far infrared (FIR) sensors. Pre-processing, including denoising and image enhancement, is necessary to extract the texture of blood vessels. The comparison is usually carried out with a skeletonized vein structure, segmented after locally adaptive image thresholding. Cross and Smith [1] developed a low cost automatic thermographic imaging system and used three vein signatures matching methods: grid-based matching, constrained sequential correlation, and fuzzy relaxation. The accuracy is estimated at 60-80% of correct recognition due to small sample size (20 people $\times$ 5 images). Lin and Fan [2] extracted dorsal hand vein pattern by using modified watershed transform based on properties of thermal images, and analyzed the dorsal palm vein patterns, to end up with EER = 2.3%. Infrared thermal hand vein patterns were also analyzed by Kumar *et al.* [3]. Their feature extraction approach based on a Gabor filter bank, enabled to achieve the false acceptance rate (FAR) of $0.1\%$ on a database of sample from 100 subjects. Personal verification system using thermal-imaged vein pattern on the back side of the hand was largely presented by Wang and Leedham [4], [5], [6]. They compared methods of acquisition of vein image using near-infrared (NIR) and far-infrared (FIR) sensors, paying attention to the advantages and disadvantages of a given solution. The study mostly concerned image acquisition, which was carried out using a thermal camera in a normal office environment. Image enhancement included denoising using a median and Gaussian low-pass filtering and image normalization to have some pre-defined mean and variance [7]. A locally adaptive thresholding was utilized to extract the vein patterns from the background. Structures extracted in this way were then skeletonized and compared using the Line Segment Hausdorff Distance. The authors report zero FAR and zero FRR on a small database.

Several methods of hand image segmentation based on thermal features were presented by Font-Aragones *et al.* [8], [9] and Faundez-Zanuy *et al.* [10]. The authors propose a segmentation method based on Fisher ratio, cluster analysis and using active shapes for both visible-light and thermal hand images. The research was carried out on images acquired from 104 subjects and the claimed accuracy is close to 85% in the case of the hand and 56% in case of the finger. Wang [11] used a thermal sensor to extract hand geometry-based features in a more accurate way when compared to using visible-light imaging only. The extension neural network (ENN) was used in classification of features and allowed to achieve 99% accuracy for a small database of samples from 30 subjects.

The only paper presenting how hand thermal data can be used in biometric recognition was proposed by Czajka and Bulwan [12]. Instead of a thermal camera, the authors used a proprietary thermal matrix of 1,000 thermal sensors delivering low-resolution heat maps. Feature selection was based on "minimum Redundancy, maximum Relevance" (mRMR) method preceded by dimensionality reduction of a feature space, for which principal component analysis (PCA) and linear discriminant analysis (LDA) were used. The authors report minimum EER = $6.67\%$ achieved for k-nearest neighbors classifier on the dataset of thermal maps acquired from $50$ subjects. It is noteworthy that the acquisition of images by the thermal matrix used in this work requires a physical contact of a hand with the thermal sensors and accurate positioning of the hand during time period of minimum $4$ seconds.

To our best knowledge, there is no research to date that proposes a biometric system based on thermal maps acquired by an infrared camera in an unconstrained environment. Also, we are not aware of any other database of thermal images of the hand that would be publicly available to researchers.

## III. Database of Thermal Hand Images

Thermal hand images used in this study were acquired by the FLIR SC645 thermal sensor [13] in a setup that did not require any hand stabilization or exact hand positioning. Each thermal sample has a resolution of $640 \times 480$ pixels and automatic focus adjustment implemented in the camera help to get sharp images. All samples were acquired in an office environment with air conditioning and the ambient temperature set to $22^{\circ}$C (thermal comfort for an average human [14]). Other factors that might influence the hand temperature, such as health condition, environment condition right before the measurement, time from last meal, etc., were recorded and they are included in the metadata.

The volunteers were asked to raise their hands up to the chest level and present their palm facing the camera. The palm plane was at the right angle to the sensor optical axis. The measurement took place without any direct contact with the sensor. During a single presentation ten images were acquired at $6.3$ FPS (frames per second). That is, a single presentation lasted approximately $1.6$ seconds. Volunteers were asked to present their hand at least five times in a single session, hence at least five series of $10$ thermal maps were acquired on each visit. To research a temporal stability of thermal maps, samples were acquired in three different sessions: two first organized on the same day, but separated by a several minutes, and the third carried out approximately a two weeks after the first one.

We have collected $21,000$ thermal maps in total from left and right hands of $70$ different subject. $26$ were males and $44$ were females. Each volunteer signed a consent and agreed to collect the thermal maps and corresponding metadata. Each record has been assigned a fake identifier making the data anonymous. This database is available for research purposes[1].

## IV. Pre-processing of Thermal Hand Images

### A. Hand Segmentation

Unconstrained acquisition calls for a precise image segmentation as the first step. This process leads to removal of unnecessary information from the image and to keep only its important properties. It is assumed that thermal images contain two classes of temperatures that should be separated: related to a hand and to a background. Two methods of setting a global binarization threshold were evaluated. The first, proposed by Otsu [15], divides pixels into two groups to maximize between-to-within-class variance. Since Otsu's method makes no assumptions on the distributions of pixel intensities, we observed correct segmentation in areas that were warmer than an ambient temperature by at least $0.2^{\circ}C$. However, hand areas with temperatures similar to the background, such as finger tips, were segmented incorrectly, as depicted in Fig. 1).

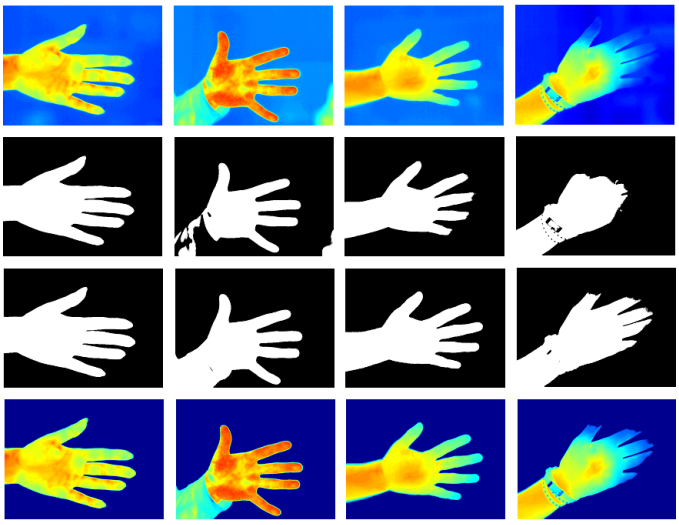

Fig. 1: **Top row**: sample thermal images. **Second row:** segmentation masks obtained by Otsu's method. **Third row**: segmentation masks obtained by GMM-based method. **Bottom row:** thermal images with GMM-based masks applied. GMM-based approach ends up with more accurate segmentation than Otsu's method.

This problem can be partially solved by using Gaussian Mixture Models (GMM) [16] to approximate distributions of hand and background pixels. GMM-based method improved image segmentation especially in areas of fingertips that often have a temperature close to ambient temperature (cf. two bottom rows in Fig. 1). However, the segmentation based on a global threshold does not work well in cases of cool fingers (see last column in Fig. 1). Hand temperature distribution depends significantly on external conditions and is closely related to location of blood vessels and dynamics of metabolic processes. The fact that the fingers are the coolest part of a hand is not only related to thermal diffusion. Our organism has priorities in maintaining adequate temperature of various body parts: the energy is first used to heat important internal organs such as brain, heart or liver, and the extremities are heated in the second order [17]. Also, incorrect segmentation was observed more frequently for female hands than these of males, possibly relating to the menstrual cycle. Thus, in this work we use only central part of the palm to extract biometric features. The exception is the CNN approach, in which we use whole-hand images for training and identity recognition.

[1]BioBase-Hand-Thermal v1.0, http://zbum.ia.pw.edu.pl/EN/node/46

### B. Region of Interest and Thermal Data Representation

To find the region of interest (ROI) used further to extract biometric features, we inscribe a circle with the largest possible radius into the hand's contour. That means, based on the binary segmentation mask, a *distance map* between all inner points and hand silhouette is first determined. Each element of the distance map is the Euclidean distance to the nearest point lying on the contour. The maximum element of the distance map determines the circular ROI (Fig. 2).

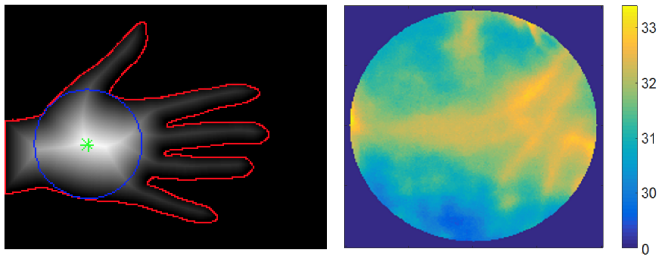


Fig. 2: **Left:** The gray level corresponds to a distance between a given point and the hand silhouette (red). The largest circle (ROI) possible to be inscribed into the hand contour is shown in blue, and its center in shown as green star. **Right:** The local temperature map within the resulting ROI.

All regions of interest must be normalized to account for varying image scale in unconstraint acquisition scenario. Experiments carried out to determine the best ROI size for BSIF-based feature extraction ended up with $101 \times 101$ pixels. Convolutional neural network applied also in this study is based on the VGG-16 architecture, forcing the input images to be upscaled to $224 \times 224$ pixels, and hence this resolution was used in a data-driven approach.

In addition, to account for the varying range of temperatures observed for different subjects and sessions, raw thermal images were normalized to the range of $\langle 0, 1 \rangle$, after which a histogram equalization operation has been applied, Fig. 3.

## V. Feature Extraction and Classification

Two approaches for feature selection and classification are proposed and compared: *feature engineering* deploying texture

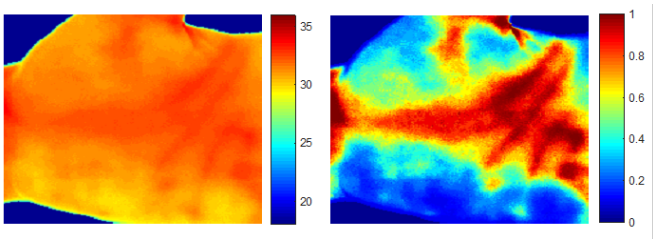


Fig. 3: Thermal image representations: raw thermal image as acquired by the camera **(right)**, thermal image normalized to $\langle 0, 1\rangle$ range with histogram equalization **(left)**.

descriptors (Binarized Statistical Image Features and Gabor wavelets) and *feature learning* based on convolutional neural networks trained in different environmental conditions. The next subsections discuss both approaches and the obtained results.

### A. Application of Texture Descriptors

*1) Binarized Statistical Image Features:* BSIF [18] gained a lot of attention in various computer vision tasks and biometrics. It generates a series of binary images that can be used as binary codes, or can be concatenated to form a grayscale image used to calculate a histogram-based image descriptor. Each bit of the output binary code is associated with a specific filter and the code length is determined by the number of filters in the filter bank. The authors suggest using from five to twelve different filters in a single filter bank. Filter kernels suggested by BSIF authors have 8 different resolutions: $3 \times 3$, $5 \times 5$, $7 \times 7$, $9 \times 9$, $11 \times 11$, $13 \times 13$, $15 \times 15$, and $17 \times 17$ pixels.

The filter kernels in BSIF approach are selected to maximize the statistical independence among filter responses. Three kinds of filters trained with different sets of images are considered in this work:

- original kernels proposed by Kannala and Rahtu [18] trained on natural images (*e.g.*, grass, stones, fur, trees, landscapes),
- kernels trained on multiple patches of twenty thermal images selected randomly from the corpus collected in this work,
- kernels trained on multiple patches of twenty images chosen randomly from a database of infrared hand vein samples [6].

Having a kernel $w_i$ and an image patch $x$ of the size equal to the size of a kernel, a local filter response is expressed as a correlation between $w_i$ and $x$ [18]:

$$s_i = \sum_{u,v} w_i(u,v)x(u,v) \tag{1}$$

A binary code is obtained by quantization of the local responses, namely: bit $b_i = 1$, when $s_i > 0$ and $b_i = 0$ otherwise. Taking into account $n$ different filter kernels selected for a filter bank, we get $n$ bits of the code for each pixel of the input image. This set of bits obtained for each pixel can be converted into a number in the range of $\langle 0; 2^n - 1\rangle$ that can be further used to calculate histogram-based image descriptor or to visualize the BSIF output as an image. Filters used in this work differed with size and design, so the essential part of this research was a selection of optimal set of kernels to extract valuable biometric information.

*2) Gabor Filters:* Gabor filters have a wide range of applications including image processing and recognition. The two-dimensional real-valued Gabor kernel is given by:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{x_\theta^2}{2\sigma_x^2}+\frac{y_\theta^2}{2\sigma_y^2}\right)} \cos\left(\frac{2\pi}{\lambda}x_\theta + \psi\right) \tag{2}$$

$$x_\theta = x\cos\theta + y\sin\theta \tag{3}$$

$$y_\theta = -x\sin\theta + y\cos\theta \tag{4}$$

where $\sigma_x$ and $\sigma_y$ define the width of the Gaussian envelope in $x$ and $y$ dimensions, $\theta$ is the kernel orientation, and $\psi$ is the phase. In all our experiments $\psi = 0$ and $\sigma_x = \sigma_y = 1$.

*3) Selection of Optimal BSIF Filter Banks and Gabor Kernels:* Both BSIF- and Gabor-based features depend on selection of filter parameters. Original BSIF filters [18] were analyzed for all sizes from $3 \times 3$ to $17 \times 17$, and BSIF filters retrained with thermal and vein images were analyzed for sizes from $3 \times 3$ to $23 \times 23$. Each BSIF-based approach is then based on a single, best-performing filter bank. A similar procedure was performed for setting the parameters of Gabor filters. 8 different orientations, namely $\theta \in \{90°, 60°, 45°, 30°, 0°, -30°, -45°, -60°, \}$ and 10 wavelengths, namely $\lambda \in \{3, 5, 7, 9, 11, 13, 15, 17, 19, 23\}$ were considered. As in BSIF, the best performing single Gabor kernel was used.

Figure 4 illustrates an example raw thermal image transformed into feature sets using four feature extraction methods evaluated in this paper.

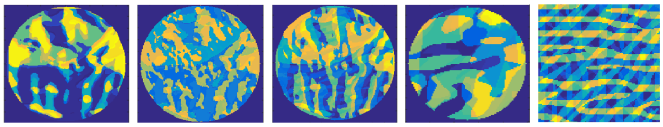

Fig. 4: Features obtained from an example thermal image. **First:** BSIF code obtained for original filters [18]. **Second:** BSIF code for filters trained on thermal hand images. **Third:** BSIF code based on filters trained on vein images. **Fourth:** encoded magnitude using Gabor filtering. **Fifth:** encoded phase using Gabor filtering.

*4) Alignment of Thermal Images:* Due to the fact that the acquisition does not require any hand stabilization and a subject can present palm in an arbitrary way, images must be aligned prior to calculation of a comparison score. Thus, in this work we use distance maps described in Sec. IV-B to align thermal images. Fig. 5 presents example distance maps of two different subjects.

These distance maps are not stable on the whole palm area. However, we observed that the central part of a line

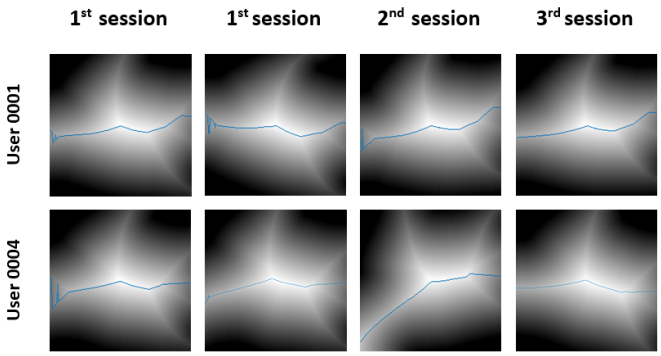


Fig. 5: Example distance maps of two subjects obtained in different sessions. Blue lines represent maximum values of the Euclidean distance from the wrist region to the middle fingertip.

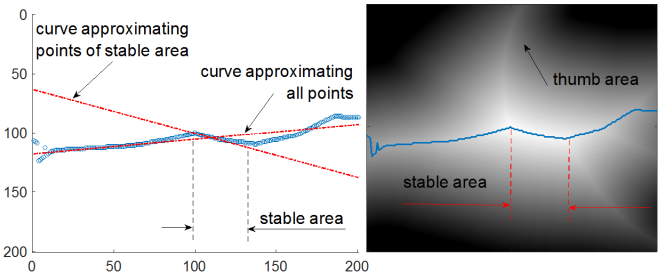


Fig. 6: Estimation of the thermal hand image rotation based on the stable part of a distance map.

representing maximum values of the Euclidean distance from the wrist region to the top of a middle finger (shown in blue in Figs. 5 and 6, and denoted as *stable area*) aligns well with hand rotation and is stable across acquisitions and subjects. The first point of the stable line segment was set in the central point of the ROI, and the distance of the second point was set experimentally to $96\%$ of the ROI radius. The line segment in a stable area was then approximated by a straight line using least mean square minimization resulting in an absolute rotation of a hand during acquisition. Consequently, the corrective image rotation was based on the difference of absolute rotations of two samples being compared.

*5) Thermal Stability Maps:* Stability of inner hand temperature may be uneven within ROI and is related to the location and pattern of the blood vessels. Since during the enrollment more than one thermal sample is collected, we can obtain the *stability maps*, *i.e.*, sections of ROI with the smallest standard deviation of thermal data for each person. These maps can then be used to calculate weight matrices for all ROI pixels:

$$w(x,y) = 1 - \frac{\sigma(x,y)}{\sigma_{\max}} \tag{5}$$

where $\sigma(x,y)$ is a standard deviation of a pixel $(x,y)$ of the thermal map, and $\sigma_{\max}$ is the maximum standard deviation observed for a given subject during enrollment. Note that $w(x,y)$ approaches 1 for very stable pixels (due to the standard deviation $\sigma(x,y)$ being close to zero), and $w(x,y) \approx 0$ for unstable areas (due to the standard deviation $\sigma(x,y)$ being close to the maximum observed value).

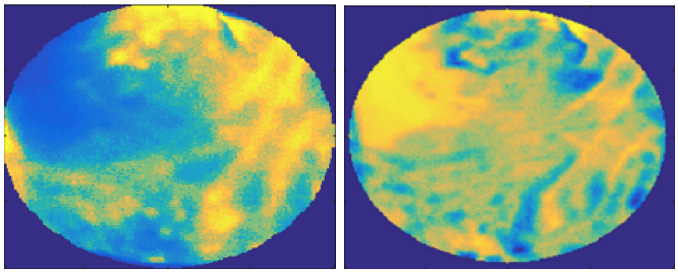

Fig. 7: Example region of interest **(left)** and a corresponding stability map $w$ **(right).**

### *B. Application of Convolutional Neural Network*

Convolutional neural networks have become widely used in computer vision problems, especially biometric recognition of faces. CNNs process images all the way from pixels to decisions owing to a hierarchy of feature extractors. Each convolutional layer extracts features from output of the previous layer, and the last, fully-connected layers perform classification. The main upside of such structures is that the convolutional kernels are learned directly from the data. Hence, there is no need to "handcraft" our feature extractors, which may be helpful in our case where our intuition related to the best hand thermal features may be incomplete or wrong. This flexibility of the CNN may be also a pitfall. The network that is insufficiently regularized by our prior knowledge about the problem may have a poor performance in cross-dataset and subject-disjoint testing.

In this study, we employed two CNN architectures for the purpose of recognizing thermal hand images:

*1) SimpleNet:* this architecture was designed and trained from scratch, and consists of $4$ convolutional layers with rectified linear units (ReLU), each of them followed by a pooling layer: max pooling after the first convolutional layer and average pooling after the remaining convolutional layers, Fig. 8. Finally, two fully connected layers with ReLU activations followed by a softmax layer provide the estimation of class membership probability for input samples. Stochastic gradient descent with momentum $m = 0.9$ was used in minimization. The initial learning rate $0.05$ has been gradually decreased to $0.0005$ in the last epochs. We have also increased the size and variability of the training set. For each ROI in the training set we added its three rotated versions: the original sample rotated by $\pm 5^\circ$ and the original sample normalized by aligning its absolute hand rotation (cf. Sec. V-A4) to the horizontal position.

*2) VGG-based CNN:* the second architecture is a well-known VGG-16 model, created by the Visual Geometry Group at the University of Oxford [19], adapted and fine-tuned using the same training set as for training the SimpleNet. It consist of 13 convolutional layers with ReLU. Every second

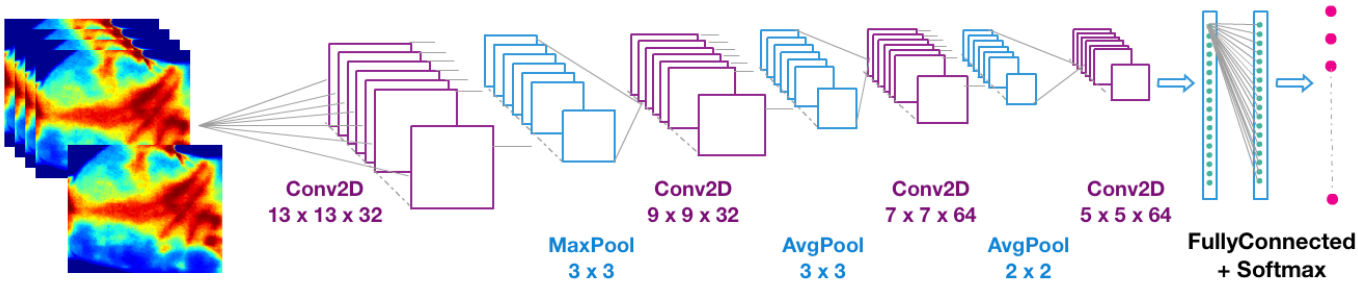


Fig. 8: Architecture of the custom convolutional neural network SimpleNet applied in this work.

convolutional layer is followed by a max pooling layer. VGG contains three fully connected layers with ReLU activations and dropout regularization. For minimization, the stochastic gradient descent with momentum $m = 0.9$ was used. The learning rate was $0.0001$. VGG is trained on natural images, but has been shown to achieve good results in numerous recognition problems, and with minor modifications and fine-tuning can be adapted to a different task that it was originally trained for.

## VI. Experiments and Results

### A. Training and Testing Subsets

Two scenarios for template creation are considered.

**First scenario:** only data acquired in the first session was used to calculate biometric references. We used $85\%$ of the first session data as the training set. Testing set consisted of $15\%$ of the data acquired in the $1^{st}$ session, $100\%$ of the data acquired in the $2^{nd}$ and $3^{rd}$ sessions. This scenario allows determining the biometric capability of thermal characteristics (intra-session comparisons) and evaluate the temporal stability of hand thermal maps assuming that the enrollment is performed during a single visit (inter-session comparisons).

**Second scenario:** we merged $85\%$ of data acquired in sessions $1$ and $2$ to compose the training set, while test was performed on $100\%$ of $3^{rd}$ session samples. This scenario allows to evaluate the method when the enrollment can be organized on a single day, yet the time gap between acquisitions is a few minutes.

We made $10$ random splits into evaluation and testing subsets, and hence $10$ independent training and testing experiments were made to calculate interval estimators of the error rates. The exception is the VGG-based approach, which was trained only on a single train/test split, due very long CNN training times. Since a single hand presentation resulted in $10$ images, and we had $5$ presentations in one session (cf. Sec. III), in each experiment we selected randomly one sample for each presentation. This allows to generate $5$ gallery images. To generate genuine and impostor comparisons, we selected randomly one probe image among all samples collected for $5$ presentations. $\chi$-square statistic was used to calculate the comparisons score between BSIF codes:

$$\chi^2 = \frac{1}{2}\sum_{i,j}\frac{\big(B(i,j) - B_0(i,j)\big)^2}{B(i,j) + B_0(i,j)} \tag{6}$$

where $B_0$ is the BSIF code of the gallery image, and $B$ is the BSIF code of the probe image. The size of the BSIF kernels achieving the best performance on the training set were:

- $17 \times 17$ for original BSIF kernels and biometric reference based on the first-session images,
- $15 \times 15$ for original BFIS kernels and biometric reference based on the first- and second-session images,
- $21 \times 21$ for all the remaining setups.

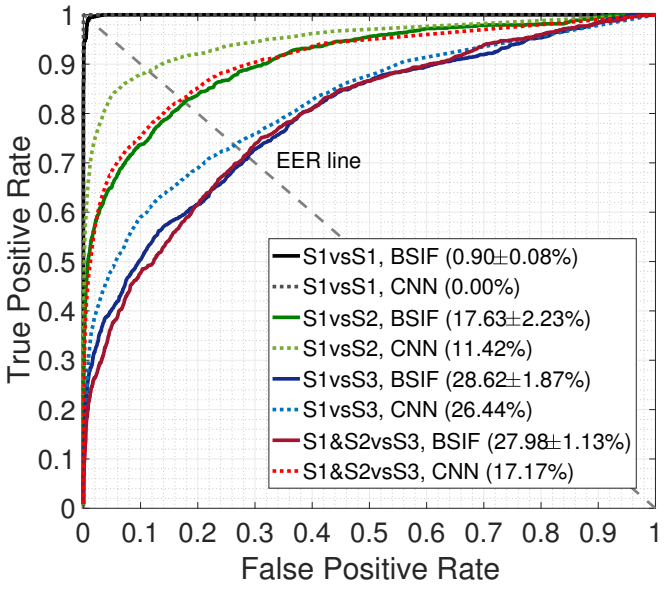


Fig. 9: ROC curves obtained for BSIF codes weighted with the stability maps **(solid lines)** and ROCs obtained for the CNN-based approach **(dotted lines)**. EER values are shown in brackets.

### B. Question 1: uniqueness of thermal hand maps

Table I presents the equal error rate values obtained from the BSIF-based approach, for two scenarios of data usage for template creation (cf. Sec. VI-A), namely employing the

| Method | training on **S1** testing on **S1** | training on **S1** testing on **S2** | training on **S1** testing on **S3** | training on **S1+S2** testing on **S3** |
|---|---|---|---|---|
| **Original BSIF** | $\mathbf{5.34 \pm 1.92\%}$ | $20.01 \pm 1.49\%$ | $\mathbf{28.76 \pm 2.06\%}$ | $\mathbf{30.24 \pm 2.81\%}$ |
| **BSIF trained with thermal data** | $8.66 \pm 2.55\%$ | $21.82 \pm 2.70\%$ | $32.79 \pm 2.94\%$ | $32.79 \pm 2.94\%$ |
| **BSIF trained with vein data** | $9.08 \pm 1.38\%$ | $\mathbf{19.15 \pm 1.78\%}$ | $31.74 \pm 2.72\%$ | $31.74 \pm 2.72\%$ |
| **Original BSIF** (with stability maps) | $0.90 \pm 0.08\%$ | $17.63 \pm 2.23\%$ | $\mathbf{28.62 \pm 1.87\%}$ | $\mathbf{27.98 \pm 1.13\%}$ |
| **BSIF trained with thermal data** (with stability maps) | $2.21 \pm 0.39\%$ | $\mathbf{17.55 \pm 2.09\%}$ | $30.93 \pm 2.69\%$ | $29.85 \pm 1.50\%$ |
| **BSIF trained with vein data** (with stability maps) | $\mathbf{0.36 \pm 0.47\%}$ | $17.86 \pm 1.75\%$ | $28.95 \pm 2.11\%$ | $28.95 \pm 2.11\%$ |
| **Gabor-based** (with stability maps) | $0.28 \pm 0.29\%$ | $22.20 \pm 1.45\%$ | $32.98 \pm 3.31\%$ | $29.98 \pm 1.43\%$ |
| **CNN-based** (SimpleNet) | $0.40 \pm 0.45\%$ | $28.80 \pm 0.57\%$ | $38.20 \pm 1.02\%$ | $36.80 \pm 1.13\%$ |
| **CNN-based** (VGG-16) | $\mathbf{0.00\%}$ | $\mathbf{11.42\%}$ | $\mathbf{26.44\%}$ | $\mathbf{17.17\%}$ |

TABLE I: Equal error rates (EERs) along with their standard deviations obtained from 10 random train-test splits for: BSIF filters trained with different data, the same BSIF filters weighted with thermal stability maps, Gabor-based representation, and classification based on a deep convolutional networks, for four different train-test scenarios.

data from the first session to generate enrollment references (columns 2-4 in Table I), and then the data from both the first and the second sessions (column 5 in Table I). The first immediate observation is that thermal hand maps may indeed deliver unique biometric features, with the lowest $ERR = 5.34\%$ for original BSIF filters (trained on natural images). Interestingly, re-training the BSIF filters with thermal hand data did not lower the error rates in most scenarios, and neither did the re-training of BSIF with palm vein data.

However, employing stability maps to appropriately weigh the unstable regions to lower their impact on the performance enabled us to further lower the intra-session EERs to below 1% (Fig. 9, black solid line). Such favorable average EERs below 1% were achieved for BSIF and for Gabor-based coding, for which stability maps were also employed. The CNN-based method (SimpleNet) also produces low ERR below 1%, when testing data comes from the same session that was employed for building the reference templates. VGG-based approach, in the same conditions, achieves perfect recognition accuracy (Fig. 9, black dotted line). **Therefore, the answer to Question 1 is affirmative: hand thermal information is unique if considered in a short time horizon and has a potential to serve as a biometric identifier**.

### C. Question 2: temporal stability of thermal hand maps

The immediate observation after analyzing the results for comparison between sessions (columns 3-5 in Table I) is that the performance drops significantly when time between acquisitions increases, see Fig. 9. The best average $EER = 27.98\%$ was obtained for original BSIF kernels when session 1 and session 2 data were fused for the training. The SimpleNet CNN seems to learn some additional session-specific properties of the data and even when the training set is enhanced with artificially rotated samples, this solution achieves $EER = 36.80\%$ in the same scenario. The VGG-based CNN seems to be the most promising solution, achieving $EER = 17.17\%$. **Thus, the answer to Question 2 is negative: hand thermal information as used in this study has limited temporal stability**.

### D. Question 3: the best approach

The results presented in Table I do not reveal an obvious winner. It seems that re-training the BSIF filters with thermal and vein images does not increase the overall accuracy when compared to the filters trained on natural images. This may suggest that the limited accuracy obtained for intra-session comparisons is related mainly to properties of hand temperature and not to the filtering kernels. The largest gain in accuracy for the BSIF-based approaches is achieved when the stability maps are used when calculating the comparison scores. The CNN-based approach allows achieving recognition rates that exceed those of the BSIF-based approach in every scenario, however, a closed-set recognition is a limitation of this method. **Hence, the answer to Question 3 is that incorporating the information about the stability of thermal image areas brings the highest increase in accuracy for the BSIF-based approaches, but employing a deep convolutional networks has a potential for even higher recognition rates**.

## VII. Conclusions

This work proposes the first known to us biometric recognition methods based on hand images acquired by thermal sensor without the requirement of hand stabilization. Processing of

thermal data acquired in unconstrained scenario requires image segmentation and alignment that is not trivial. This paper shows that distributions of hand temperatures are unique and with the proposed processing pipeline the recognition can be close to perfect. On the other hand, we present that temporal stability of hand thermal maps is limited.

Unconstrained environment assumed in this work makes this a very convenient and hygienic system. Also, thermal information is difficult to copy and present to the thermal sensor in a way that would be difficult to detect. This allows effective presentation attack detection without the need to employ other hardware or software. Hence, the next necessary step in this research must be related to appropriate modeling of intra-session variability that include both biomedical and technical aspects to propose processing methods that are more robust to large variations of hand temperature in time.

To facilitate research in this area, and to follow recommendations on reproducible research, the database of thermal images collected for this study is made available to all interested researchers for non-commercial purposes.

## ACKNOWLEDGMENTS

The authors would like to cordially thank the members of the Biometrics Scientific Club at the Warsaw University of Technology for their help with building the database used in this study. The authors would like to thank Mateusz Trokielewicz for his remarks on the initial draft of this paper. Also, the authors would like to thank NVIDIA for supporting them with a GPU unit that was used in experiments.

## REFERENCES

[1] J. Cross and C. Smith, "Thermographic imaging of the subcutaneous vascular network of the back of the hand for biometric identification," *Institute of Electrical and Electronics Engineers 29th Annual 1995 International Carnahan Conference on Security Technology*, pp. 20–35, 1995.

[2] C.-L. Lin and K.-C. Fan, "Biometric verification using thermal images of palm-dorsa vein patterns," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 199–213, Feb 2004.

[3] A. Kumar, M. Hanmandlu, V. K. Madasu, and B. C. Lovell, "Biometric authentication based on infrared thermal hand vein patterns," in *2009 Digital Image Computing: Techniques and Applications*, Dec 2009, pp. 331–338.

[4] L. Wang and G. Leedham, "A Thermal Hand Vein Pattern Verification System," *Image Analysis. ICAPR 2005. Lecture Notes in Computer Science, vol 3687. Springer, Berlin, Heidelberg*, 2005.

[5] ——, "Near- and Far- Infrared Imaging for Vein Pattern Biometrics," *Video and Signal Based Surveillance, 2006. AVSS '06. IEEE International Conference on Sydney, Australia, Australia*, 2006.

[6] L. Wang, G. Leedham, and S. y. Cho, "Infrared imaging of hand vein patterns for biometric purposes," *Institution of Engineering and Technology (IET)*, vol. 1, pp. 113–122, 2007.

[7] L. Hong, Y. Wan, and A. Jain, "Fingerprint Image Enhancement: Algorithm And Performance Evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 777–789, 1998.

[8] X. Font-Aragones, M. Faundez-Zanuy, and J. Mekyska, "Thermal hand image segmentation for biometric recognition," *Security Technology (ICCST), 2011 IEEE International Carnahan Conference, Barcelona, Spain*, 2011.

[9] ——, "Thermal hand image segmentation for biometric recognition," *IEEE Aerospace and Electronic Systems Magazine*, vol. 28, no. 6, pp. 4–14, June 2013.

[10] M. Faundez-Zanuy, J. Mekyska, and X. Font-Aragones, "A new hand image database simultaneously acquired in visible, near-infrared and thermal spectrums," *Cognitive Computation*, vol. 6, no. 2, p. 230240, 2013.

[11] M.-H. Wang, "Hand recognition using thermal image and extension neural network," *Mathematical Problems in Engineering*, vol. 2012, pp. 1–15, 2012.

[12] A. Czajka and P. Bulwan, "Biometric verification based on hand thermal images," in *2013 International Conference on Biometrics (ICB)*, June 2013, pp. 1–6.

[13] "FLIR, User manual FLIR A645 S.C.25," [last accessed: Feb. 20, 2016].

[14] K. Fabbri, "A Brief History of Thermal Comfort: From Effective Temperature to Adaptive Thermal Comfort," *Indoor Thermal Comfort Perception*, 2015.

[15] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan 1979.

[16] Z.-K. Huangac and K.-W. Chaub, "A new image thresholding method based on gaussian mixture model," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 899–907, 2008.

[17] N. A. Taylor, C. Machado-Moreira, A. van den Heuvel, J. Caldwel, E. A. Taylor, and M. J. Tipton, "The roles of hands and feet in temperature regulation in hot and cold environments," in *Proceedings of the 13th International Conference on Environmental Ergonomics*, Aug 2009, pp. 405–409.

[18] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov 2012, pp. 1363–1366.

[19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *https://arxiv.org/abs/1409.1556*, 2014.