IEEE Xplore: https://ieeexplore.ieee.org/abstract/document/9210996
GI Digital Library: https://dl.gi.de/handle/20.500.12116/34345

Improved Liveness Detection in Dorsal Hand Vein Videos using Photoplethysmography

Johannes Schuiki ♦ Andreas Uhl

Department of Computer Sciences, University of Salzburg, Austria

Abstract

In this study, a previously published infrared finger vein liveness detection scheme is tested for its applicability on dorsal hand vein videos. A custom database consisting of five different types of presentation attacks recorded with transillumination as well as reflected light illumination is examined. Additionally, two different methods for liveness detection are presented in this work. All methods described employ the concept of generating a signal through the change in average pixel illumination, which is referred to as Photoplethysmography. Feature vectors in order to classify a given video sequence are generated using spectral analysis of the time series. Experimental results show the effectiveness of the proposed methods.

Contents

This project was partly funded from the FFG KIRAS project AUTFingerATM under grant No. 864785 and the FWF project "Advanced Methods and Applications for Fingervein Recognition" under grant No. P 32201-NBL.

1 Introduction

The human body provides a great number of biometric traits which due to their inter-personal uniqueness provide a high level of distinctiveness. Especially identification systems using vascular patterns, that is, structures in blood vessels, have become an important field of research in the last decades. Two important properties of these biometric traits are that they are a) not left behind like fingerprints or DNA and b) usually hidden inside the human body which makes the use of specially designed imaging devices almost inevitable. Common parts of the body for gathering recordings of these vessels are hands (including wrists and fingers) and eyes. This research focuses on information captured from the dorsal view of hand vessels. Such systems make use of the fact that the hemoglobin in the human vessels have a relatively high molar extinction coefficient, which is a measurement for attenuation of electromagnetic waves with respect to a certain wavelength, in the near-infrared (NIR) range. In general one differs between two types of illumination variants (not counting the combination of both techniques into a hybrid version): *Transillumination*, where the hand is placed inbetween the light source and the imaging sensor and *Reflected light*, where the NIR light comes from the same direction as the camera. By doing so, and as the name suggests, the reflected light is then again captured by the imaging sensor. An illustration of these two illumination variants is given in fig. 1 and examples of the resulting images can be seen in fig. 2.

Biometric identification systems in general suffer from what is known as presentation attacks [1]. Here, data is presented to the biometric capturing subsystem with the goal of interfering (e.g. for impersonation) with this system. Countermeasures are therefore referred to as presentation attack detection (PAD). Usually, PAD methods try to classify the data presented to the capturing subsystem into either attack presentations or bona fide presentations. It has been shown also for the modalities used in this work that presentation attacks exist [2], [3] and are a potential threat that has to be dealt with. In particular these presentation attacks are conducted by either printing previously acquired bona fide images on paper or utilizing a digital display (e.g. from a smartphone) which is then presented to the imaging sensor. The act of ensuring that the extracted features descent from actual blood vessels is referred to as liveness detection and builds a subset of PAD methods. Related work proposing countermeasures for presentation attacks on NIR vein identification systems can be split into the following categories:

Multimodal approaches, that achieve liveness detection through additional hardware or by looking at more than one trait [4]–[7]. Other related works provide algorithmic contributions, where authors try to distinguish real and spoof data by software. *Deep learning based* approaches that employ convolutional neural networks (CNNs) build one of those branches. Another idea for PAD is to utilize differences in *quality*, *skin texture* and *spatial frequency components* in still images. A broad overview of the still image approaches (including CNN, quality, texture and frequency) is given in table 14.1 from [8].

A different approach to accomplish PAD is to use *video sequences* that utilize differences of features in adjacent frames. The idea proposed by [9] applies a technique called Eulerian motion magnification on transillumination finger vein videos. This method amplifies minor changes caused by active blood flow. Recently [10] analyzed this scheme for a custom dorsal hand vein dataset that was captured with transillumination as well. The following methods build upon a common pre-processing step that allows using the video data for temporal analysis of blood flow in the human hand. With every frame, either on single pixel level, on a region of interest (ROI), or with the image as a whole (for the latter two the average pixel illumination is calculated), a time series is generated. The series should resemble the characteristics of a beating heart, indicating life. This is a form of Photoplethysmography which is explained in section 3. In 2008, [11] used this form of temporal analysis on a ROI of the human hand with reflected light imaging using two different wavelengths, namely 660 and 880nm. The work already contained the observation that one dominant peak in the frequency spectrum around 1.4 hertz would most likely be the heart rate. In [12] the goal was actually to present a new method for palm vein extraction using reflected light imaging, however the procedure includes the estimation of the power spectral density over an 8 minute video sequence. Again, a dominant peak around 1.4 hertz was reported. Another technique for discrimination of bona fide from attack videos was proposed in [13]. Here, the average minima and maxima from the time series form features for classification, but it is also shown that this method can be spoofed by a blinking LED that imitates pulse. In [14] again the similarity to blood pressure signals is shown with videos from a custom built NIR transillumination finger vein capturing device, but lacks information on how to use this information for liveness detection.

The work in this paper analyses the applicability of

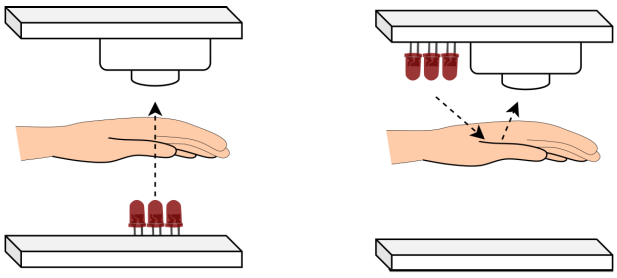

Figure 1: Modes of operation of the capturing device: Transillumination (left) and Reflected Light (right).

a finger vein liveness detection scheme proposed by Bok et al. [15] on a custom dorsal hand vein dataset explained in section 2. While the reference work uses a database which only consists of sequences captured with transillumination, here we also evaluate the results on reflected light data. Furthermore, this work proposes two additional methods for the construction of a feature vector in order to separate bona fide from presentation attack video sequences. Section 3 explains the existing work and the contributions made by this paper. Section 4 describes the evaluation methods and contains a discussion of the results.

2 Video Database

The imaging installation used to capture the video sequences is similar to the one that was used in [10], [16]. The imaging sensor is a Canon EOS 5D MarkII DSLR with removed IR blocking filter, that is placed on the top side of a wooden box. Additionally, a 850mn pass-through filter is attached to the scanner. The installation is able to capture data in two illumination variants, namely transillumination & reflected light. In order to accomplish that, a NIR surveillance lamp including 50 940nm NIR LEDs is positioned on the bottom of the capturing device, while 6 950nm LEDs are attached on the top side of the box. Figure 1 illustrates the two modes of operation. The database contains both hands from 13 participants, that were captured with both lighting versions, resulting in 52 genuine video sequences.

Every video has five presentation attack counterparts, namely a printed frame on cardboard (in this work referred to as *Paper*), the same printed frame but moving back and forth approximately with a pace of 1 hertz to simulate heartbeat artefacts (*Paper Moving*), one frame displayed on a smartphone (*Smartphone*), a frame displayed on a smartphone with a programmed sinusoidal translation (*Smartphone Moving*) and the frame shown on a smartphone with a rhythmic zoom-in-zoom-out effect (*Smartphone Zooming*). Both variants of motion, i.e. translation and scaling, are meant to simulate a heartbeat-like variation of illumination on the dorsum of the hand. For the smartphone attacks a custom android application was created to apply the transformations. Therefore, in total the whole database consists of 312 video sequences, all of resolution 1920x1080. For generation of the attacks, only a region of interest (ROI) was selected. Fig. 2 shows examples of genuine and attack frames in both illumination variants. The sequences are of varying length. Bona fide videos range from 14.75 to 25.29 seconds, attack sequences from 10.41 to 21.49 seconds. Every sample was captured with a constant frame rate of 29.97 frames per second.

3 Remote NIR Photoplethysmography using Spectral Analysis

Plethysmography is the act of measuring changes in volume in different areas of the body. Examples for such measurements are limb circumference, impedance (electric resistance) on certain bodyparts or the amount of air that is being exhaled to conclude lung capacity. One variant of Plethysmography is called Photoplethysmography (PPG) and refers to analysis of optical signals acquired through imaging techniques. PPG makes use of the observation that heart beats generate measureable changes in the human body resulting in a periodical pattern that repeats every cardiac cycle. This non-invasive method is used in devices such as pulse oxymeters or for heart rate estimation in sport watches. In the case of NIR imaging, one utilizes the fact that the oxygen saturation in blood has its peak directly after the heart beat and reaches a minimum shortly before the next impulse. Wei et al. [17] have shown that such blood pressure measurements do not only contain the heart rate, but due to their shape characteristics also include harmonics (i.e. integer multiples of a dominant fundamental frequency) that can be mathematically modelled through exponential decay. This paper exploits this observation for the construction of a feature vector that indicates real blood flow.

A recent publication on presentation attack detection for finger vein image sequences has shown that information from PPG analysis can be used as classification criteria [15]. The proposed approach employs the concept of generating a one dimensional time series by calculation of the average pixel brightness in every frame of the infrared finger vein video. The time series is first zero padded to reach

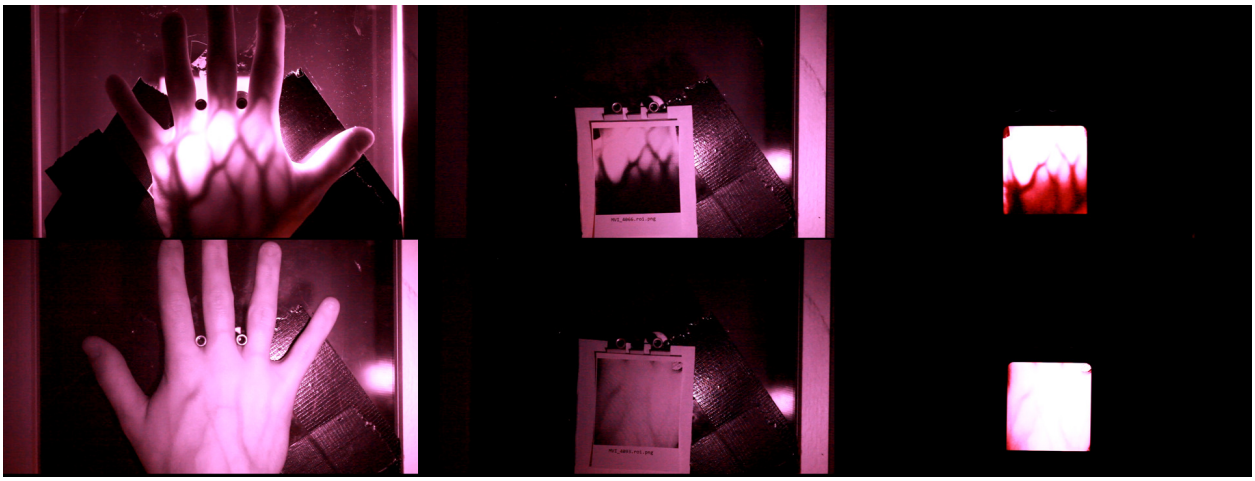

Figure 2: Example frames from NIR videos; top row: transillumination, bottom row: reflected light; left: real video frame, middle column: paper spoof and right: smartphone spoof, respectively

a higher frequency resolution and then transformed into Fourier space using the discrete Fourier transform (DFT). Next, frequency components less than 1.0 and higher than 3.0 hertz are dropped. By taking the magnitude of every frequency component in this range, using a spacing of 0.04 hertz, a 50 dimensional feature vector (FV) is constructed. For classification a support vector machine (SVM) with a radial basis function kernel was used.

3.1 Proposed Methods

Two different ways for FV construction are explained in this section. Both build upon a common basis. An overview is given in figure 3. The methods apply for both illumination variants.

Baseline: Due to the fact that the custom imaging device has two screwed pins where the middle finger has to be placed inbetween, every hand is placed on the same predetermined spot. The pins can also be seen on the subplots of figure 2. Therefore, as a first step, a fixed 600x600 ROI is cropped such that only the dorsal view of the hand (from knuckles to wrist) remains. Similar to the reference method, a time series is built containing values of average pixel illumination (8 bit grey value) from this ROI. In order to center the time series around the x axis, the mean value (DC component) is subtracted. A rectangular window with a size of 150 frames is applied and shifted over the time series with a stepsize of 1 frame. For detrending and removal of other artefacts, an infinite impulse response high-pass filter with a steepness of 0.95 and a cutoff frequency $f_{cut} = 0.5\ Hz$ is applied to every window. As a final step for the baseline, a zero padding of 3000 zeros is attached to the windowed and filtered signal before transformed to frequency space using DFT. The highest frequency and resolution of the spectrum is given by $f_{max} < \frac{29.97}{2} = 14.985\ Hz$ and $\Delta f = \frac{29.97}{150+3000} = 0.0095\ Hz$.

Method 1: For every window, the global $argmax$ (i.e. the frequency where the spectrum has its highest peak) is stored in a temporary list. From that list, a histogram is generated with a bin size of 0.05 Hz ranging from 0.5 to 2 Hz covering all windows. Values outside of that range do not contribute to the histogram. The histogram is normalized since video sequences are of varying length and some windows may have its peak outside the range. The FV for method 1 is given by that normalized histogram and depicts a form of majority voting per window.

Method 2: We define our global $argmax(\mathcal{F})$ as f_{HR} and the $max(\mathcal{F})$ as m_{HR}, where $\mathcal{F}$ is the frequency spectrum ignoring phase information. If f_{HR} is the heart rate, then due to [17] one would expect local maxima at the harmonic frequencies (integer multiples of f_{HR}) with a certain magnitude as well. Therefore, for the first 4 harmonics (i.e. $i * f_{HR}$, $i \in \{2...5\}$) a search window of $\pm\frac{f_{HR}}{5}$ is defined. In addition to f_{HR} and m_{HR}, we temporarily store the local argmax and corresponding maxima as the a quotient w.r.t the global argmax & maxima as f_i and m_i. The final FV is constructed by taking the means and medians (Md) for all stored values over every window, i.e.:

$$(\overline{m_{HR}}, \overline{m_1}...\overline{m_4}, Md(m_{HR}), Md(m_1)...Md(m_4),$$
$$\overline{f_{HR}}, \overline{f_1}...\overline{f_4}, Md(f_{HR}), Md(f_1)...Md(f_4))$$

Exceeding f_{max} with a search window counts as 0

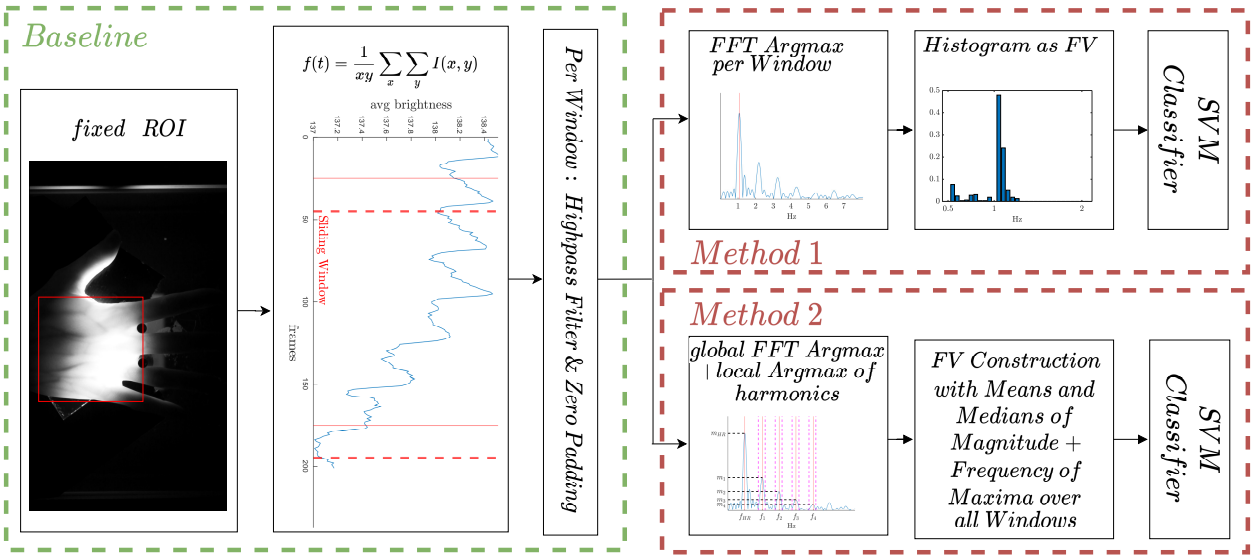

Figure 3: Block diagram of the proposed methods.

for the entries in question. This case, by construction, can only occur if $f_{HR} * 5 + \frac{f_{HR}}{5} \geq f_{max} \Rightarrow f_{HR} \geq 2.88\ Hz$, which is out of range for a reasonable heart rate.

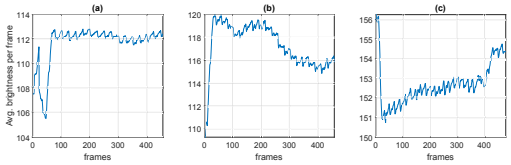

Figure 4: Types of errors: ascending/descending trend, steps, peaks & fluctuation.

4 Results

Bok et al. [15] used a SVM classifier with a radial basis kernel function and parameters $C = 10, \gamma = 0.001$. Matlabs *fitcsvm* function was used for classification. Therefore the parameters were set as: 'BoxConstraint' = 10 and 'KernelScale' = $\frac{1}{\sqrt{\gamma}}$. In addition, a simple linear kernel was used for classification. The data under test was split into training and evaluation data using the Leave One Out Cross Validation (LOOCV) principle. Results are reported in compliance with the ISO/IEC 30107-3:2017 standard [18], which defines metrics for presentation attack detection such as Attack Presentation Classification Error Rate (APCER) and Bona Fide Presentation Classification Error Rate (BPCER). Table 4 contains the results of both test scenarios. For the proposed methods, the amount of data was given as 26 videos per presentation attack and bona fide respectively. Therefore, the error rate step is $\frac{1}{26} = 3.85\%$. Since for the Bok et al. approach data was split into chunks, between 92 and 109 sequences per category were available for training a classifier. One can see that, with few exceptions, the proposed method 2 is superior to the other two methods under test using the linear kernel classifier. Even with the settings from Bok et al., except for paper based spoofs with reflected light illumination, method 2 yields acceptable results. In comparison to finger vein videos, where the imaging sensor is relatively close to the finger, hand vein imaging offers a lot more possibilities for unwanted errors. As depicted in example time series in figure 4, even slight movements of the hand have noticeable effects in the pixel-averaged time series. That is the reason why using the whole video together with high-pass filtering yields very robust results in comparison to just cutting the sequence into chunks as proposed in [15].

Another measure to circumvent inconsistencies in the time series is given by the windowing. The step size was set to one frame in order to assign more weight on spots where the pulse is clearly visible, which is the case for the major part of the signal. Unfortunately, the proposed approaches are more costly in terms of computational resources which also reflects in the processing time. On average, processing a single video took 32.564 seconds using method 1 and 32.778 seconds using method 2. Compared to the approach from the reference work, where processing all 312 videos (video-slicing included) took

Table 1: The upper table shows the SVM results with a RBF kernel, BoxConstraint of 10 and a γ value of 0.001 as proposed in [15]; The second table contains results with a simple linear kernel; best results are highlighted **bold**.

Spoof Method		Bok et al.		Method 1		Method 2	
RBF: C=10, $\gamma = 0.001$		APCER	BPCER	APCER	BPCER	APCER	BPCER
Transill.	Paper	**10.31**	30.77	11.54	15.38	11.54	**3.85**
	Paper Moving	**4.35**	37.50	26.92	11.54	11.54	**0.00**
	Smartphone	**0.00**	89.42	3.85	7.69	**0.00**	**0.00**
	Smartphone Mov.	77.36	89.42	**0.00**	19.23	19.23	**0.00**
	Smartphone Zoom	59.43	17.31	**3.85**	19.23	11.54	**0.00**
Refl. Light	Paper	**0.00**	80.37	53.85	61.54	7.69	**23.08**
	Paper Moving	49.06	5.61	**23.08**	**3.85**	42.31	7.69
	Smartphone	63.55	6.54	7.69	**3.85**	**3.85**	**3.85**
	Smartphone Mov.	23.08	6.54	3.85	**0.00**	**0.00**	3.85
	Smartphone Zoom	15.24	7.48	3.85	**0.00**	**0.00**	3.85
linear		APCER	BPCER	APCER	BPCER	APCER	BPCER
Transill.	Paper	11.34	18.27	**7.69**	26.92	11.54	**0.00**
	Paper Moving	**7.61**	16.35	26.92	19.23	11.54	**7.69**
	Smartphone	13.21	54.81	7.69	26.92	**0.00**	**0.00**
	Smartphone Mov.	17.92	57.69	11.54	11.54	**3.85**	**0.00**
	Smartphone Zoom	27.36	40.38	7.69	11.54	**0.00**	**0.00**
Refl. Light	Paper	12.84	71.03	61.54	53.85	**0.00**	**3.85**
	Paper Moving	25.47	**1.87**	15.38	3.85	**3.85**	7.69
	Smartphone	46.73	6.54	7.69	19.23	**3.85**	**3.85**
	Smartphone Mov.	28.85	6.54	3.85	**3.85**	**0.00**	**3.85**
	Smartphone Zoom	18.10	5.61	**3.85**	**0.00**	**3.85**	3.85

246 milliseconds, this is a downside of the introduced methods. The time consumption is caused by the fact that we apply a fourier transform together with Matlabs highpass filter function on every window with a step size of one frame. The average time consumption per window was 113 milliseconds. As processing hardware, an AMD Ryzen 5 1600X six-core processor was used for all experiments.

5 Conclusion

In this work, the applicability for a finger vein liveness detection scheme was tested on a custom dorsal hand vein dataset in two illumination variants. Two additional classification methods are described. One constitutes a windowed majority voting for the human heart rate by looking for a maximum in the frequency spectrum. The other method exploits the shape characteristic of a typical blood pressure measurement by looking around multiples of the estimated heart rate for other spectral peaks, which has not been used for liveness detection so far. Doing so, the method delivers a certain invariance to the value of the actual pulse and has proven to be a fairly robust approach for both, transillumination and reflected light video sequences. Although the methods are not very efficient in terms of computational cost yet, they form a starting point for a robust unimodal approach for presentation attack detection in dorsal hand vein videos.

References

[1] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics systems under spoofing attack: An evaluation methodology and lessons learned", *IEEE Signal Processing Magazine*, vol. 32, no. 5, 2015.

[2] I. Patil, S. Bhilare, and V. Kanhangad, "Assessing vulnerability of dorsal hand-vein verification system to spoofing attacks using smartphone camera", in *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2016.

[3] P. Tome and S. Marcel, "On the vulnerability of palm vein recognition to spoofing attacks", in *2015 International Conference on Biometrics (ICB)*, 2015.

[4] A. Krishnan, T. Thomas, G. R. Nayar, and S. Sasilekha Mohan, "Liveness detection in finger vein imaging device using plethysmographic signals", in *Intelligent Human Computer Interaction*, Springer International Publishing, 2018.

[5] S. Crisan and B. Tebrean, "Low cost, high quality vein pattern recognition device with liveness detection. workflow and implementations", *Measurement*, vol. 108, 2017.

[6] S. Crisan, B. Tebrean, and T. E. Crisan, "Multimodal liveness detection system for hand vein biometrics", in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2018.

[7] M. K. Shahin, A. M. Badawi, and M. E. Rasmy, "A multimodal hand vein, hand geometry, and fingerprint prototype design for high security biometrics", in *2008 Cairo International Biomedical Engineering Conference*, 2008.

[8] J. Kolberg, M. Gomez-Barrero, S. Venkatesh, R. Ramachandra, and C. Busch, "Presentation attack detection for finger recognition", in *Handbook of Vascular Biometrics*. Springer International Publishing, 2020.

[9] R. Raghavendra, M. Avinash, S. Marcel, and C. Busch, "Finger vein liveness detection using motion magnification", in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2015.

[10] T. Herzog and A. Uhl, "Analysing a vein liveness detection scheme", in *Proceedings of the 8th International Workshop on Biometrics and Forensics (IWBF'20)*, 2020.

[11] J. Zheng, S. Hu, V. Azorin Peris, A. Echiadis, V. Chouliaras, and R. Summers, "Remote simultaneous dual wavelength imaging photoplethysmography: A further step towards 3-d mapping of skin blood microcirculation", *Proc SPIE*, vol. 6850, 2008.

[12] H. Zhang and D. Hu, "A novel preprocessing method for palm vein", *Advanced Materials Research*, vol. 658, 2013.

[13] H. Ding, "Anti-spoofing a finger vascular recognition device with pulse detection", in *24th Twente Student Conference on IT*, University of Twente, 2015.

[14] J. H. Han, J. Kim, and E. C. Lee, "Single-camera vision-based vein biometric authentication and heart rate monitoring via infrared imaging analysis", in *Advances in Computer Science and Ubiquitous Computing*, Springer Singapore, 2018.

[15] J. Bok, K. Suh, and E. C. Lee, "Detecting fake finger-vein data using remote photoplethysmography", *Electronics*, vol. 8, 2019.

[16] A. Gruschina, "Veinplus: A transillumination and reflection-based hand vein database", in *Proceedings of the 39th annual workshop of the Austrian association for pattern recognition (OAGM'15)*, 2015. arXiv: 1505.06769 [cs.CV].

[17] C.-C. Wei, C.-M. Huang, and Y.-T. Liao, "The exponential decay characteristic of the spectral distribution of blood pressure wave in radial artery", *Computers in Biology and Medicine*, vol. 39, no. 5, 2009.

[18] ISO, "Information technology — biometric presentation attack detection — part 3: Testing and reporting", International Organization for Standardization, ISO ISO/IEC 30107-3:2017, 2017.