Rotation Tolerant Finger Vein Recognition using CNNs

1st Bernhard Prommegger*
Department of Computer Sciences
University of Salzburg
5020 Salzburg, AUSTRIA
bprommeg@cs.sbg.ac.at

2nd Georg Wimmer*
Department of Computer Sciences
University of Salzburg
5020 Salzburg, AUSTRIA
gwimmer@cs.sbg.ac.at

3rd Andreas Uhl
Department of Computer Sciences
University of Salzburg
5020 Salzburg, AUSTRIA
uhl@cs.sbg.ac.at

Abstract—Finger vein recognition deals with the recognition of subjects based on their venous pattern within the fingers. The majority of the available systems acquire the vein pattern using only a single camera. Such systems are susceptible to misplacements of the finger during acquisition, in particular longitudinal finger rotation poses a severe problem. Besides some hardware based approaches that try to avoid the misplacement in the first place, there are several software based solutions to counter fight longitudinal finger rotation. All of them use classical hand-crafted features. This work presents a novel approach to make CNNs robust to longitudinal finger rotation by training CNNs using finger vein images from varying perspectives.

Index Terms—Finger vein recognition, longitudinal finger rotation, rotation tolerance, CNN

I. INTRODUCTION

The performance of finger vein recognition systems suffers from environmental conditions (e.g. temperature and humidity) and deformations due to misplacement of the finger, typically including shifts, tilt, bending and longitudinal rotation. The influence of some of these misplacements can be reduced or even prevented completely either during acquisition by adding support structures for finger positioning or a correction during pre-processing, feature extraction or comparison. Especially longitudinal finger rotation is hard to avoid. In [1], the authors showed that existing finger vein data sets contain longitudinal rotation to a non neglectable extend. By eliminating only longitudinal finger rotation (all other condition remain unchanged), they achieved performance increases of up to 350%. This indicates that longitudinal finger rotation is not only a problem in selected use cases, but a general problem in finger vein recognition. As finger vein recognition systems evolve towards contact less acquisition (e.g. [2], [3]), problems due to finger misplacements will become more severe.

Longitudinal finger rotation is hard to counteract as it changes the positioning of the veins and their visibility due to a non-linear transformation. As can be seen in Fig. 1, the acquired vein pattern of a finger differs depending on its rotation. There is already some work on rotation detection and compensation for single-camera systems. Prommegger *et*

al. [5] analysed different approaches and showed that existing recognition systems, even when applying rotation compensation, cannot handle rotational distances of >30°. Others try to tackle the problem by acquiring the vein pattern from different perspectives (e.g. [6], [7]). The disadvantage of multi-camera systems are the increased cost and complexity.

Recently, finger vein recognition systems using convolutional neural networks (CNN) are getting more attention. These systems are either not designed to counter fight longitudinal rotation (e.g. [8], [9]) or require the acquisition of finger vein images from multiple perspectives [7]. Therefore, this paper is the first to present a CNN based rotation tolerant single camera finger vein recognition system. The proposed idea is to train CNNs using finger vein images from varying perspectives. There are two different sources for these images: (1) images that were actually taken from different perspectives and (2) augmented images using a novel approach that simulates longitudinal rotations during training. This way, the CNNs should learn to recognize the connection between images that come from different angles and thus to recognize the non-linear distortion caused by the rotation.

II. CNN ARCHITECTURES

To demonstrate that our proposed approach to increase the rotation tolerance of CNNs is independent of the used CNN architecture and loss function, two different CNN architectures and loss functions are used in our experiments:

Squeeze-Net (SqNet) with triplet loss function: The advantage of the triplet loss compared to more common loss functions (e.g. SoftMax) is that the CNNs learn to group images of the same classes together in the CNN feature output space and separate images from different classes, instead of directly classifying images. So, contrary to common loss functions, CNNs can also be applied to images whose classes are not included in the training data. This property is crucial in biometric applications. The triplet loss using the squared Euclidean distance is defined as follows:

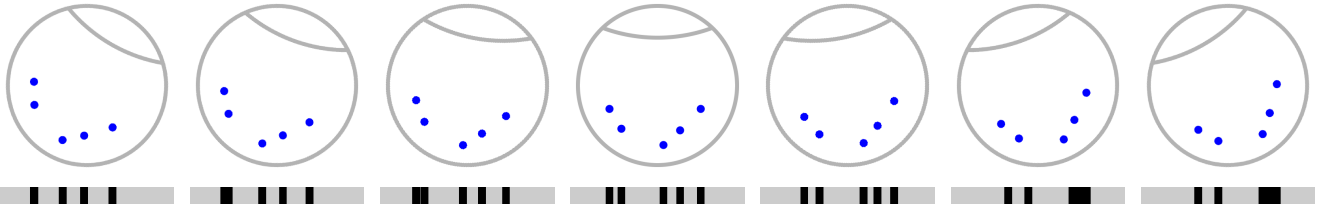$$L(A, P, N) = \max(||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + \alpha, 0) \quad (1)$$

Fig. 1: Schematic finger cross section showing five veins (blue dots) rotated from -30° (left) to +30° (right) in 10° steps. The projection (bottom row) of the vein pattern changes depending on the rotation angle according to a non-linear transformation (originally published in [4])

where A and P are two images from the same class (finger), N from a different one. α is a margin that is enforced between positive and negative pairs (in our case $\alpha = 1$), and $f(x)$ is the CNN output of an input image x. Same as in [9], we employ hard triplet selection and the Squeeze-Net architecture.

DenseNet with SoftMax loss: A more common approach than using the triplet loss is to train a net with the common Soft-Max loss function and then use the net as feature extractor for evaluation by using the CNN activations of intermediate layers. This approach has already been applied in prior work (e.g. [8]). As network architecture we employ the DenseNet-161. For evaluation, we remove the final layer and thereby get a 2208 dimensional feature vector output when feeding an image through the network.

III. Training Data

As already described, the rotated versions of the training images are provided using two different approaches. In the first approach, the images are acquired at different rotation angles, while in the second approach the rotation is artificially generated using data augmentation. Using vein images that were actually captured from different perspectives for CNN training is of course more effort than generating the rotated versions with the help of data augmentation. It should be noted that rotated samples of the same finger are only needed for training. The actual angles of rotation of these samples do not necessarily have to be known, as long as the acquired samples cover the rotational range for which the recognition system should be tolerant. This can be achieved by e.g. placing the finger in different rotations on the existing single camera capturing device or by rotating the camera and illumination module around the finger as done for the employed data sets. This is certainly a reasonable expense for commercial products. Recognition is still applied using vein images from a single perspective.

All images for training and evaluation are normalized using *Circular Pattern Normalization* (CPN) [5]. In principle CPN corresponds to a rolling of the finger surface assuming a circular finger shape. After this unrolling, longitudinal rotations correspond to shifts in the acquired images.

Finger Vein Images Captured from Different Perspectives: For the training of the CNNs, finger vein images acquired from different perspectives are used (the finger is

Fig. 2: ROI of sample images of the PLUSVein-FR after applying CPN. Left: palmar view (0°), middle: vein image captured at 45°, right: 45° artificially rotated version of the palmar image.

rotated around its longitudinal axis). All images of a finger, regardless of the angle at which they were taken, are considered as the same class. As a result of this, the CNN should learn to recognize finger vein images independent of their perspective. The left and middle image in Fig. 2 show two such input images. The left image has been acquired from the palmar view (0°), the middle one from 45°. It is clearly visible that the vein pattern is vertically shifted and deformed in a non-linear manner due to the rotational difference.

CNN Training using Augmentation of Finger Vein Images: The augmented training data is generated from images acquired at the palmar view (0°). The height of a CPN image is $h_{CPN} = r \cdot \pi$, which is half the fingers perimeter with an assumed radius of r. The displacement (in pixels) that must be applied for a rotation of a defined rotation angle φ_{rotate} can be calculated as:

$$h_{shift}(\varphi_{rotate}) = \frac{2 \cdot r \cdot \pi \cdot \varphi_{rotate}}{360°} = \frac{h_{CPN} \cdot \varphi_{rotate}}{180°} \quad (2)$$

For data augmentation in the rotational range of $\pm\varphi$, the height h of the input images is enlarged by twice the maximum shift $h = h_{CPN} + 2 \cdot h_{shift}(\varphi)$. Augmentation is applied by randomly cropping patches with height h_{CPN} (and original width) of the enlarged images, which corresponds to rotations in the range of $\pm\varphi$. The right image in Fig. 2 is an artificially rotated version (45°) of the original image at the left.

IV. Experiments

Datasets: The datasets used for the experiments are the *PLUSVein Finger Rotation Dataset* (PLUSVein-FR) [10] and the *PROTECT Multimodal Dataset* (PMMDB) [11]. Both datasets provide finger vein images acquired all around the finger (360° in steps of 1°) and have been acquired using the same sensor and the same acquisition protocol. In this work, only the perspectives in the range of $\pm45°$ around the palmar view are used. The PLUSVein-FR provides vein images from 63 different subjects with 4 fingers per subject and each
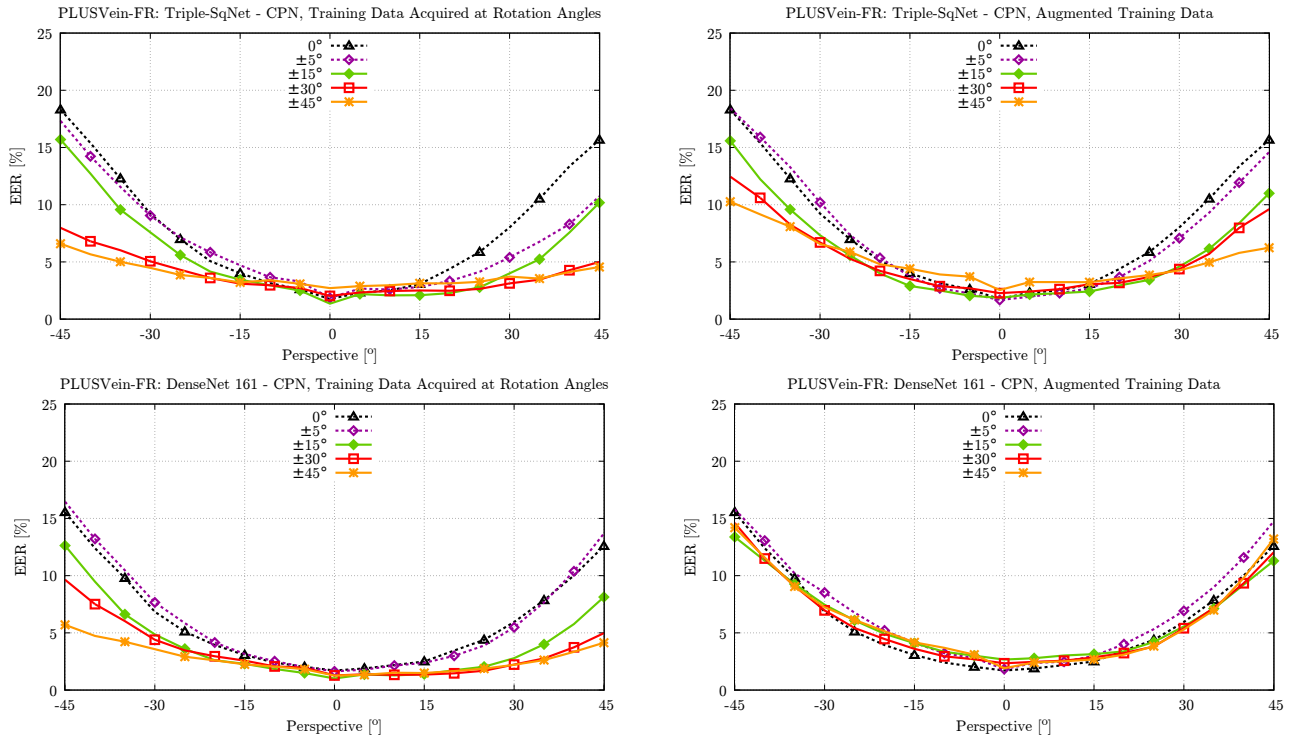
Fig. 3: Trend of the EER across different longitudinal rotations applying Triplet-SqNet and DenseNet-161 trained with different rotational ranges

finger is acquired 5 times per perspective, resulting in 1.260 finger vein images per perspective and 115.920 vein images in total. PMMDB is acquired from 29 subjects with 4 fingers per subject and either 5 or 10 images per finger (one or two sessions) for each perspective with a total of 102.987 finger vein images.

For *finger region detection* and *finger alignment*, an implementation that is based on [12] is used. The *ROI extraction* differs from [12]: Instead of cutting out a defined rectangle within the finger, the width of the finger is stretched to a fixed width and normalized using CPN. It is worth to note, that no image enhancement techniques (e.g. contrast enhancement) have been applied to the input images.

CNN Training: In order to study the influence of using training data from different longitudinal rotations on the rotation invariance of the CNNs, the range of rotation from which the training samples were taken was varied. The ranges are $0°$ (which corresponds to the training of a classical single-camera recognition system with images from palmar view) and $\pm 5°$, $\pm 15°$, $\pm 30°$ and $\pm 45°$ from the palmar view ($0°$). All experiments are executed using (1) images acquired at different rotations and (2) augmented images simulating different rotations for CNN training.

Training is performed for 400 epochs starting with a learning rate of 0.001 for SqNet and 0.005 for DenseNet. The learning rate is divided by 10 each 120 epochs. For both nets, training is performed on batches of 128 images. The images are resized to 224×224 pixels and normalized

to zero mean and unit variance before feeding them into the CNNs. The two employed nets are pre-trained on the ImageNet database. In order to ensure a 100% separation of the training and evaluation data set, the training data was taken from the PMMDB, while for evaluation it was taken from PLUSVein-FR.

Evaluation Protocol: The EER is used to assess the recognition performance. The evaluation follows the test protocol of the FVC2004 [13]. The employed similarity metric to measure the similarity between CNN feature outputs of different images (genuine and impostor scores) is derived from the Euclidean distance. To transform the Euclidean distance to a similarity metric, the Euclidean distances are inversed ($d \rightarrow 1/d$) and normalized so that the resulting similarity values range from zero to one.

Evaluation of the CNN's Rotation Invariance: For the evaluation of the rotation invariance of the CNNs, we apply networks trained using images from different rotational ranges ($0°$, $\pm 5°$, $\pm 15°$, $\pm 30°$ and $\pm 45°$). For the evaluations, the vein images acquired at a certain rotation angle φ are compared to the ones acquired at the palmar view. φ is varied from -45° to 45°.

The trend of the EERs for Triplet-SqNet are shown in the top row of Figure 3. The left image holds the results for the experiments using training images actually acquired at different angles, whereas the right plot depicts the results for the augmented training images, where the images have been acquired at the palmar view and the rotation has been

simulated as described in Section III. The plots reveal that the proposed approach to train CNNs with vein images from different rotations works quite well. For Triplet-SqNet, the recognition rates of the reference evaluation (training only with images of the palmar view) drop rapidly for increasing rotational differences. With an increasing rotational range of the actually acquired training data, this decline becomes far less pronounced. For a training range of ±45°, the EER at the palmar view (0°) is approximately 3%. For the perspectives at +45° and -45° it is still around 6% for using training data acquired at different rotation angles. Training the CNN with augmented image data improves the results as well, but not to the same extent. For the training range of ±45°, this results in EERs below 10% at +45° and -45°.

The same evaluations have been executed for DenseNet-161 (bottom row of Figure 3). Training the DenseNet-161 using images acquired from larger rotational ranges improves the recognition results, and therefore also the CNNs invariance to longitudinal rotations. Training with larger rotational ranges leads to slightly smaller improvements compared to the reference setting (training images only taken from the palmar view) as for Triplet-SqNet, but the performance of the reference settings is also noticeable better for DenseNet-161 over the whole range of ±45°. The results using augmented input data for DenseNet-161 show no clear improvements.

V. DISCUSSIONS

In order to be able to quantify the performance of the proposed method, the best performing methods for actually acquired rotated training data (DenseNet-161 using CPN and ±45°) and for using augmented training data (Triplet-SqNet using CPN and ±45°) are compared to the best performing methods of a previous evaluation on longitudinal rotations in finger vein recognition [5]. The comparison methods comprise *Principal Curvature* (PC) [14], *Maximum Curvature* (MC) [15], *Deformation Tolerant Feature-Point Matching* (DTFPM) [16], a SIFT based approach [17] and *Finger Vein Recognition With Anatomy Structure Analysis* (ASAVE) [18]. The evaluations in [5] showed, that the mentioned recognition schemes achieve their best results when combined with *Elliptic Pattern Normalization* (EPN) [19] and the *Fixed Angle Approach* [5] to counteract longitudinal finger rotation.

The results presented in the left plot of Fig. 4 indicate, that the performance of classical hand-crafted recognition systems is good if the samples contain little to no longitudinal rotations. With an increasing rotational distance between the probe and enrolment samples, the performance drops noticeable. The best performing classical systems are the simple vein pattern based approaches PC and MC. For more sophisticated approaches (DTFPM, SIFT and ASAVE), the absolute performance degradation due to longitudinal rotation is higher. In contrast, the EER of the proposed CNN based approaches are higher for smaller rotations, but the drop of the performance is lower when the rotation increases. The right plot of Fig. 4 should visualize this effect by plotting the relative performance degradation (RPD), calculated as RPD $= \frac{ERR_{rotated} - ERR_{palmar}}{ERR_{palmar}}$,
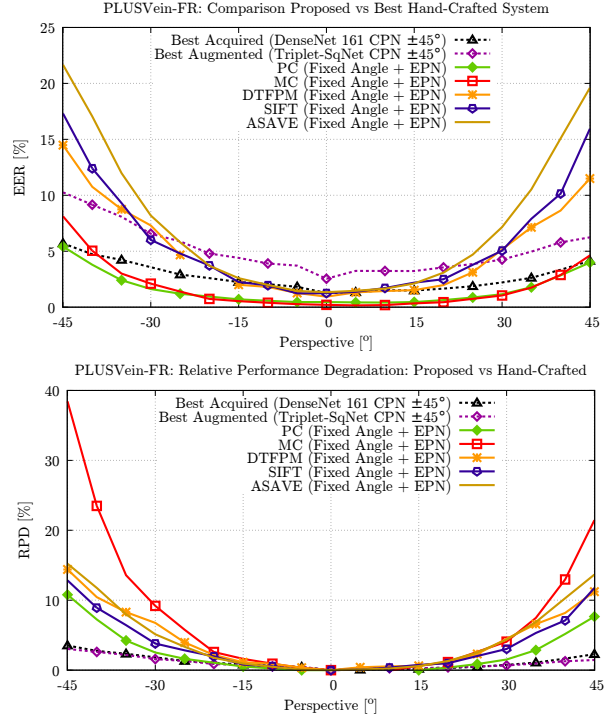

Fig. 4: Trend of the EER (left) and RPD (right) across different longitudinal rotations comparing the proposed systems with hand-crafted single perspective recognition systems

of the different methods. It is obvious that the two CNNs using our proposed training strategies are most robust against longitudinal rotation as their drop in performance is the least.

Besides to the robustness against longitudinal rotation, the proposed CNN approach has some additional advantages over traditional hand-crafted solutions:

Pre-Processing: Most traditional finger vein recognition systems require different pre-processing steps (e.g. image enhancement) that have to be tailored to each data set. Apart from the ROI extraction, the approach presented in this article does not require any pre-processing except of resizing and normalization (which are standard preprocessing steps for CNNs and require hardly any computation time and do not require any adaption to different data sets).

Cost of Time: Once the CNNs are trained, executing a single comparison is very fast. On average, feature extraction takes 7 ms, a single comparison 0.01 ms. This is way faster as for hand-crafted approaches applying time consuming approaches to increase rotation invariance. Experiments in [5] have shown that e.g. for PC with the rotation compensation scheme "fixed angle" and EPN feature extraction takes just below 130 ms, and a comparison 2.4 ms.

VI. CONCLUSIONS

In this article, we presented a novel CNN training strategy to increase the CNN's tolerance against longitudinal finger rotation. It is the first CNN-based approach to achieve rotation tolerance on single camera finger vein recognition systems

and it can be applied to any CNN, regardless of the used net architecture and loss function. We showed, that by training the CNNs using vein images acquired from different perspectives, the tolerance with respect to longitudinal finger rotation of the CNNs can be increased noticeable. For Triplet-SqNet, the same holds true if images acquired from a single perspective are artificially rotated into different perspectives for the training (data augmentation), but to a smaller extent.

Although the trained CNNs do not yet achieve the same baseline performance (when all samples are acquired from the same perspectives) as systems utilizing classic hand-crafted features, their tolerance against longitudinal finger rotation is exceptional good. The performance degradation caused by longitudinal finger rotation is noticeable lower for the trained CNNs compared to classical systems. Besides the CNN's robustness to rotations, other advantages compared to classical systems are that CNNs do not need any special pre-processing (besides of the ROI extraction) and that a single biometric comparison is very fast.

References

[1] B. Prommegger, C. Kauba, and A. Uhl, "On the extent of longitudinal finger rotation in publicly available finger vein data sets," in *Proceedings of the 12th IAPR/IEEE International Conference on Biometrics (ICB'19)*, Crete, Greece, 2019, pp. 1–8.

[2] Y. Matsuda, N. Miura, Y. Nonomura, A. Nagasaka, and T. Miyatake, "Walkthrough-style multi-finger vein authentication," in *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE'17)*, 2017, pp. 438–441.

[3] C. Kauba, B. Prommegger, and A. Uhl, "Combined fully contactless finger and hand vein capturing device with a corresponding dataset," *Sensors*, vol. 19(22), no. 5014, pp. 1–25, 2019.

[4] Bernhard Prommegger, Christof Kauba, and Andreas Uhl, "Longitudinal finger rotation - problems and effects in finger-vein recognition," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG'18)*, Darmstadt, Germany, 2018.

[5] B. Prommegger, C. Kauba, M. Linortner, and A. Uhl, "Longitudinal finger rotation - deformation detection and correction," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 2, pp. 123–138, 2019.

[6] B. Prommegger and A. Uhl, "Rotation invariant finger vein recognition," in *Proceedings of the IEEE 10th International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, Tampa, Florida, USA, 2019.

[7] W. Kang, H. Liu, W. Luo, and F. Deng, "Study of a full-view 3D finger vein verification technique," *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2019.

[8] H. G. Hong, M. B. Lee, and K. R. Park, "Convolutional neural network-based finger-vein recognition using NIR image sensors," *Sensors*, vol. 17, no. 6, pp. 1297, 2017.

[9] G. Wimmer, B. Prommegger, and A. Uhl, "Finger vein recognition and intra-subject similarity evaluation of finger veins using the cnn triplet loss," in *Proceedings of the 25th International Conference on Pattern Recognition (ICPR)*, 2020, pp. 1–7.

[10] B. Prommegger, C. Kauba, and A. Uhl, "Multi-perspective finger-vein biometrics," in *Proceedings of the IEEE 9th International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, Los Angeles, California, USA, 2018.

[11] C. Galdi, J. Boyle, L. Chen, V. Chiesa, L. Debiasi, J-L Dugelay, J. Ferryman, A. Grudzien, C. Kauba, S. Kirchgasser, M. Kowalski, M. Linortner, P. Maik, K. Michon, L. Patino, B. Prommegger, A. F. Sequeira, L. Szklarski, and A. Uhl, "Protect: Pervasive and user focused biometrics border project," *IET Biometrics*, September 2020.

[12] Y. Lu, S. J. Xie, S. Yoon, J. Yang, and D. S. Park, "Robust finger vein ROI localization based on flexible segmentation," *Sensors*, vol. 13, no. 11, pp. 14339–14366, 2013.

[13] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "FVC2004: Third Fingerprint Verification Competition," in *ICBA*. 2004, vol. 3072 of *LNCS*, pp. 1–7, Springer Verlag.

[14] J. H. Choi, W. Song, T. Kim, S-R Lee, and H. C. Kim, "Finger vein extraction using gradient normalization and principal curvature," in *Image Processing: Machine Vision Applications II*, 2009, vol. 7251, pp. 1–9.

[15] N. Miura, A. Nagasaka, and T. Miyatake, "Extraction of finger-vein patterns using maximum curvature points in image profiles," *IEICE transactions on information and systems*, vol. 90, no. 8, pp. 1185–1194, 2007.

[16] Y. Matsuda, N. Miura, A. Nagasaka, H. Kiyomiu, and T. Miyatake, "Finger-vein authentication based on deformation-tolerant feature-point matching," *Machine Vision and Applications*, vol. 27, no. 2, pp. 237–250, 2016.

[17] H. Qin, L. Qin, L. Xue, X. He, C. Yu, and X. Liang, "Finger-vein verification based on multi-features fusion," *Sensors*, vol. 13, no. 11, pp. 15048–15067, Nov 2013.

[18] L. Yang, G. Yang, Y. Yin, and X. Xi, "Finger vein recognition with anatomy structure analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2017.

[19] B. Huang, Y. Dai, R. Li, D. Tang, and W. Li, "Finger-vein authentication based on wide line detector and pattern normalization," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 1269–1272.