Algorithms for a novel touchless bimodal palm biometric system

Olegs Nikisins, Teodors Eglitis, Mihails Pudzs, Modris Greitans
Institute of Electronics and Computer Science
14 Dzerbenes Street, Riga, LV1006, Latvia
{Olegs.Nikisins, Teodors.Eglitis, Mihails.Pudzs, Modris.Greitans}@edi.lv

Abstract

The paper introduces the combination of algorithms for possibly the first bimodal biometric system capable of touch-less capturing of two biometric parameters, palm veins and palm creases, synchronously with a single image sensor. The architecture of the proposed system is based on the Detection, Alignment and Recognition pipeline. The ROI detection and alignment stages are simplified with efficient combination of hardware (lighting sources) and software. A new feature descriptor, namely Histogram of Vectors is proposed in the recognition stage. Since the capturing of images requires special conditions, the database including images of 100 individuals and ground-truth data is introduced. The analysis of performance of the system utilizes the database leading to detailed understanding of the error propagation in the automatic recognition pipeline.

1. Introduction

A growing interest in biometric systems capable of user authentication without physical contact (touchless systems) can be observed in modern palm biometrics. These systems are hygienic, compact, and more convenient to use than contact systems. However, due to additional degrees of freedom in palm placement, the accuracy of touch-less systems is decreased. These systems face many challenging tasks. Firstly, the presence of a palm in the field of view must be detected (palm detection). Secondly, region of interest (ROI) must be localized, because a palm can be captured under different orientations and distances when no palm fixing stand is used. To increase precision, more than one biometric feature is often used [16, 2, 11]. When features of two modalities are used (*e.g.*, palm veins and creases), systems are referred to as bimodal. Bimodal approaches allow to avoid typical limitations of unimodal systems, like vulnerability to spoofing attacks and low level of security. These features must be acquired and processed properly. We mention various solutions to these problems and propose working combination of algorithms for touch-less bimodal palm biometric system.

An extracted ROI implies presence of a palm in the image. Therefore, in the literature ROI selection is addressed more frequently than palm detection. For the same reason, we analyze only the ROI selection methods, but in section 3 both – palm detection and ROI extraction algorithms – are proposed. In contact systems a region of fixed location and fixed scale in the input image can serve as ROI. In touch-less systems more complicated algorithms are necessary for ROI selection due to possible variations in locations, scales and rotations. Palm contour and/or finger valley (finger-gap) points [7, 17] can serve as key-points for ROI selection. In this case ROI is usually rectangular. However, to reduce errors caused by incorrect determination of palm angle, palm ROI can also be expressed using polar-like coordinates [10]. To select ROI we use finger gaps as key-points because they can be detected in an efficient way and can also be used to determine whether the image contains the palm. We also use free-shaped ROI that maximizes area of the palm used in the extraction of the biometric information.

Feature processing is an important topic that should be discussed in detail. In outlined biometric system, each biometric feature is acquired as image. Near-infrared (NIR) and visible light (VL) images of the palm are correlated since they represent the same region of the palm. It is clear that each such image always contains some combination of palm features of different modalities, only in different proportions. For example, when capturing palm in the VL, input image contains information about creases, ridge structure, skin pigmentation, geometry, and even vein pattern (slightly visible); whereas, palm images captured in the NIR spectrum depict palm geometry, vein structure and crease pattern (slightly visible). There are several approaches to feature processing. The most widespread approach is to process both images (VL and NIR) directly [2, 11, 9], neglecting appearance of features of other modalities in the input signal. Otherwise, features of different modalities can be either combined together [16], either separated [6] before processing. A process of combining feature information from different sources into a single image before pro-
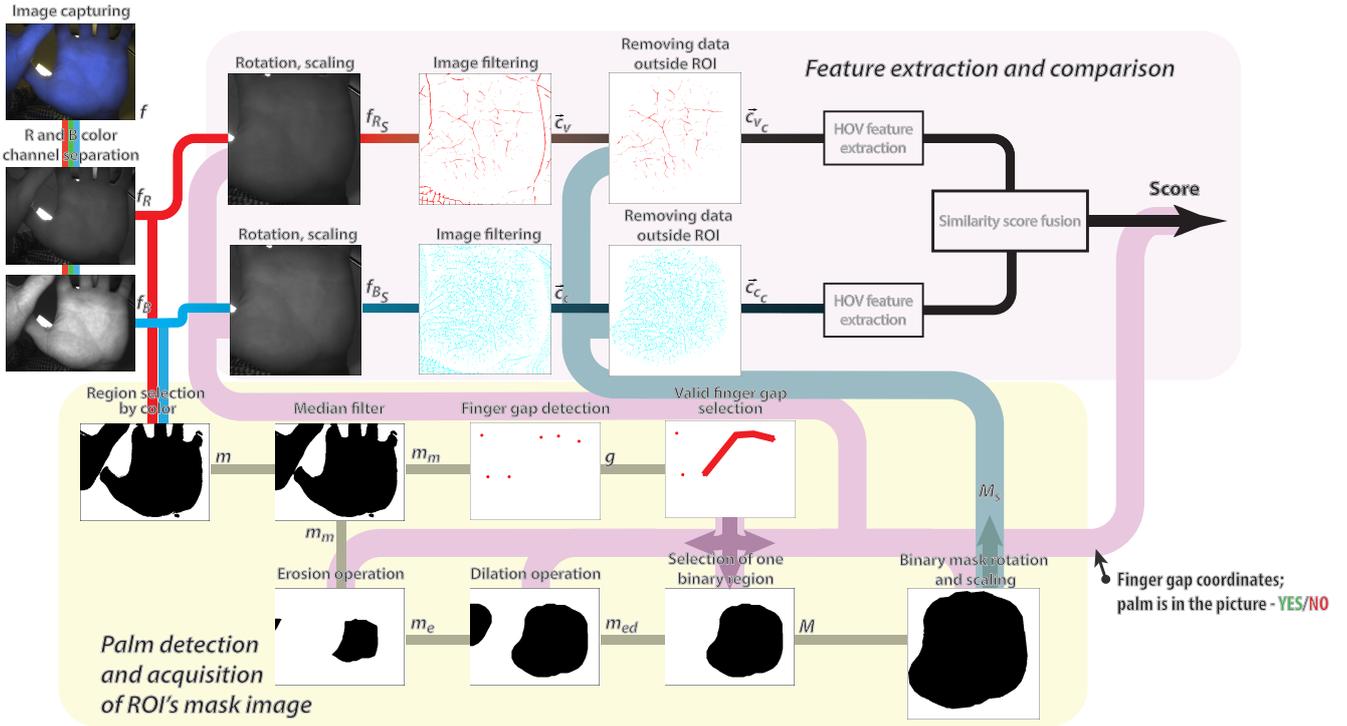
Figure 1. Architecture of the proposed system.

cessing is called sensor-level fusion. In [16] sensor-level fusion improved recognition accuracy because it increased the amount of detectable features (junction points). All mentioned approaches work because any repeatable pattern, acquired from a person's palm, can serve as biometric feature even if it depicts multiple modalities [8]. However, we think that the benefit from processing combinations of features of different modalities depends on the feature extraction and comparison approach. We further show that if histogram-based methods, like proposed algorithm Histograms of Vectors (HOV), are used, separation of vein and crease features before processing can show an improved recognition results, compared to direct image processing.

We develop and test our described algorithms in an FPGA-based system, shown in Figure 2.

The architecture of the proposed system is based on the Detection, Alignment and Recognition pipeline [14] and is shown in Figure 1. Image capturing module that is used to captures single color image and acquire two grayscale input images for further processing, is described in Section 2. Section 3 covers palm detection and ROI extraction module that is used to acquire palm finger gap coordinates and ROI binary mask. Feature extraction and comparison module incorporates an image filter (algorithms for direct image processing and modality separation are discussed) and HOV for feature extraction and further processing — Section 4. In section 5, performance of different ROI selection and fea-

ture processing approaches is analyzed using our collected database of 2000 images (100 persons, 20 images per person).

2. Image capturing principle

We utilize the technique of bimodal biometric feature acquisition using a single RGB image [6]. To ensure that palm features of different modalities – veins and creases – appear more expressed in the appropriate color channels – red (R) and blue (B) – the palm is illuminated with specific dichromatic light (a combination of blue and near-infrared (850nm) light with specific intensities) during image acquisition. The advantage of this approach is that information about both features is acquired simultaneously. This allows us to keep palm illumination constant and to stream images for the processing at the sensor's framerate. As a consequence, the feature extraction and authentication time depend solely on implemented feature processing algorithm providing rapid authentication.

3. Palm detection and acquisition of ROI's mask image

Palm can be placed at various angles and distances. Scales of observed palm features and filter masks should be matched, and HOV algorithm divides processed image into a grid with cells of fixed size, therefore, ROI should be

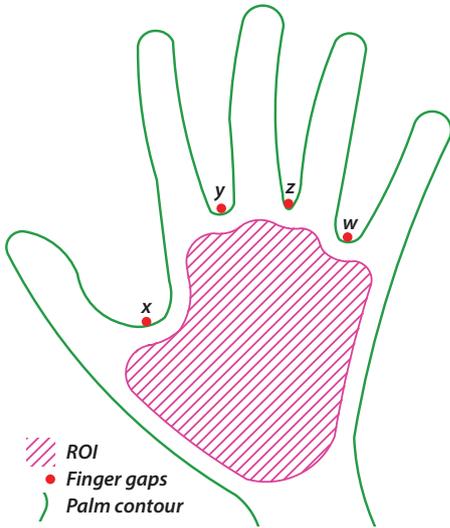Figure 2. Picture of the developed FPGA-based palm biometric system prototype

ROI
Finger gaps
Palm contour

Figure 3. Hand contour, marked fingergaps and ROI

properly rotated and scaled by means of image processing before feature extraction. Thus, additionally to selection of palm's ROI, its rotation angle and scale are determined. To maximize the amount of used biometric information we use maximum palm area with repeatable features as ROI, Figure 3. ROI excludes hand edges and parts near fingers, because these regions can be affected by ambient light and shadows. To determine whether there is a palm in the image, and to calculate ROI parameters, fingergaps (marked in Figure 3 as x, y, z and w) must be detected. A block diagram of proposed algorithm is shown in Figure 1. Our palm detection algorithms are handcrafted. To select parameters for our algorithms, we have created a small training database of multiple person palm images. To simulate normal authentication conditions palms were placed under different angles

and distances and under different background illuminations (both - indoors and outdoors).

Palm detection begins with selection of palm skin region. In the acquired images, palm skin region can be selected by color because used dichromatic light (see Section 2) appears in certain color when it is reflected from the palm [6]. Variations of palm intensity in the acquired images can be reduced (leading to increased accuracy of palm detection) by forcing image sensor to determine the exposure using only center of the image, where palm appears most frequently. Region selection by color is done using only R and B color channels as only they are employed in biometric feature extraction; in such way the resources for algorithm implementation in embedded system are reduced.

The selection of the ROI mask (Figure 1, m) can be achieved with simple threshold operations. This conclusion is based on the analysis of the colors and intensities of palm pixels and background pixels in the training database. It is assumed that input image pixels do not represent palm if the following properties are not met:

$$0.85 \cdot b > r \wedge 0.1 < r < 0.98 \wedge 0.25 < b < 1, \quad (1)$$

where r, $b \in [0, 1]$ are R and B color intensities. Such selection can be used with most background illuminations, except outdoors under a direct midday sun. To smooth the mask's m contours of obtained region for finger gap detection and to fill-in missing interior pixels, binary median filter is used (Figure 1, m_m).

It is considered that finger gaps can be located in regions, where contour of the found mask in m_m (Figure 1) forms a certain $\smile$ - shape. Such regions can be detected using a simple morphological image filter, which masks are shown in Figure 4. Result that is obtained by processing input image m_m with each finger gap mask (g_{left} and g_{right}) is calculated as follows:

$$g_{\text{left}} = \left\{ z | (m_m)_{-z} \cap N_{\text{left}} = \varnothing \wedge (m_m^c)_{-z} \cap P_{\text{left}} = \varnothing \right\} \quad (2)$$

$$g_{\text{right}} = \left\{ z | (m_m)_{-z} \cap N_{\text{right}} = \varnothing \wedge (m_m^c)_{-z} \cap P_{\text{right}} = \varnothing \right\} \quad (3)$$

The combined result, representing finger gap regions as groups of points, is calculated as $g = g_{\text{left}} \cap g_{\text{right}}$ (Figure 1). For each group that exists in the acquired binary image g only lowest middle point is left for further processing. Finger gaps of different sizes can be detected because finger-gap detection masks are sparse-like.

Not all found points represent finger gaps, *e.g.* notice that in Figure 1, g 6 $\smile$ - shape objects were found, whereas only 4 of them are finger gaps. In order to determine which of the found points denote finger gaps that belong to the palm (thus, identifying whether there is a palm in the image) found points are classified. To adjust parameters of
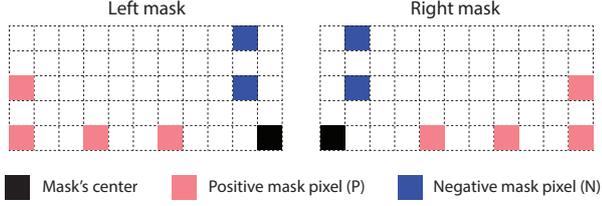
Figure 4. Masks used for finger gap detection

the classifier, we first labeled all finger gaps in the training database. Afterwards, we searched for a robust and easy to calculate parameters for valid point selection. It was determined that squared distances between every pair of points, as well as slopes of the lines connecting each pair of points provide the desired results. Additionally, we determined the range of values for selected parameters.
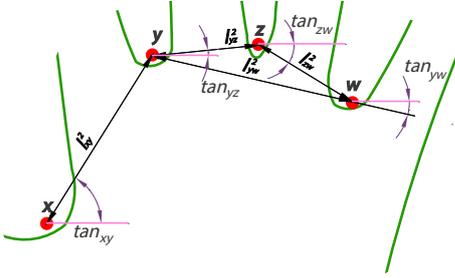

Figure 5. Parameters used in finger gap classification

Unless a group of 4 points with satisfying characteristics is found, it is assumed that the input image does not contain a palm and a new image is acquired and processed. Once the presence of a palm is confirmed, 4 fingergap points are associated with palms binary region. Before feature extraction, the acquired palm's region requires proper scaling and rotation. The scale factor is determined by assuming that the distance between y and w finger gaps should be K pixels (Figure 3). If the distance between points y and w in the acquired unscaled image is d pixels, the image should be scaled by a factor of K/d. The rotation angle is determined by assuming that the line connecting the points y and w should be horizontal.

ROI (Figure 3) should exclude palm contour and finger regions. We achieve this by performing morphological operations, erosion and dilation, on smoothed palm's mask image m_m (Figure 1).

To ensure rotation invariance, circle is used as a structuring element; to achieve scale invariance, the size of the structuring element (circle's diameter) is associated with finger gap distance d. An erosion operation is performed using circle with diameter d as structuring element (Figure 1, m_e), and is followed by dilation operation with a circle of a smaller diameter (*e.g.*, $0.8 \cdot d$, Figure 1, m_{ed}). Afterwards,

one binary region is selected using previously obtained information about valid finger gap coordinates. The remaining region is the binary mask (Figure 1, M) determining the ROI in the input image.

Now, both R and B color channels of the input image (images Figure 1, f_r and f_b), as well as the mask M of the ROI can be transformed into a new coordinate system S using affine transformations, in order to construct images with correct rotation angle and normalized scale of the palm. Afterwards, transformed R and B color channel images (Figure 1, f_{r_S} and f_{b_S} respectively) are filtered to extract biometric information, obtaining complex images $\vec{f}_v$ and $\vec{f}_c$ (Figure 1). Transformed ROI mask M_S is used to remove the image data located outside the ROI, thus producing biometric data images $\vec{f}_{v_c}$ and $\vec{f}_{c_c}$, which are fed to separate HOV feature extractors (Figure 1).

4. Feature extraction

Feature extraction can be divided into two parts – primary part, when biometric information is extracted from the R and B color channels of the transformed input image (f_{R_S} and f_{B_S}) as vector fields ($\vec{c}_v$ and $\vec{c}_c$), and secondary part, when acquired vector fields are described using HOV (Figure 1).

4.1. Primary feature extraction

By using finger gap coordinate information, transformed R and B color channel images – f_{R_S} and f_{B_S} – are acquired. It is possible to extract line-like objects (LLOs) from these images and obtain vector field representation of each image. We compare two different approaches of biometric data processing. Therefore, image filters of two types — NH-CMF for general-purpose LLO extraction (direct image processing) [12] and filters for vein and crease separation [6] — are discussed and compared.

NH-CMF This filter responds to any LLO (which include lines and gradients) of a certain width and intensity, producing complex 2D field $\vec{c}(x, y)$ representation. A line-extraction kernel $M(x, y)$ is rotated by 8 different angles φ_n, uniformly distributed in the range of $[0, \pi)$. Thus, $n = 0...7$ and $\varphi_n = \frac{n \times 180}{8}$ deg. For simplicity, general case is observed here and images of both color channels (f_{R_S} and f_{B_S}) are referred as f_f. At each kernel's rotation angle, the scalar image $f_f(x, y)$ is convolved with the rotated kernel and the following correlation value is calculated for each image point (x_0, y_0):

$$c_{f_n}(x_0, y_0) = R\left[\sum_{x,y} f_f(x, y) \cdot M(x - x_0, y - y_0; \varphi_{f_n})\right],$$

(4)

where $R[x] = \frac{x+|x|}{2}$; followed by:

$$\vec{c}_f(x_0, y_0) = \sum_n c_{f_n}(x_0, y_0) \cdot \exp(2i \times \varphi_{f_n}), \quad (5)$$

where i is the imaginary unit. Each response is a vector $\vec{c}_f(x, y)$ that characterizes underlying LLO in two ways:

- its magnitude $|\vec{c}_f(x, y)|$ represents the similarity of the neighborhood of the pixel $f_f(x, y)$ to LLO;

- its angle $\arg_f[\vec{c}_f(x, y)]/2$ represents the angle of the characterized LLO.

As $\vec{c}_f$ is calculated for each color channel, filter result is $\vec{c}_v$ and $\vec{c}_c$. Images of both color channels are filtered using kernels with different sizes, chosen to match either veins, or creases more accurately. Apart from differentiating by scale, filter is unable to segregate modalities that form the particular extracted lines or gradients. Therefore, information about both modalities appears in both resulting vector fields – $\vec{c}_{v_c}$ and $\vec{c}_{c_c}$ – only in different proportions.

Vein and Crease Filter (V-CF) These filters were developed to produce 2D vector fields that contain information about veins (vein filter, producing $\vec{c}_v$) and palm creases (crease filter, producing $\vec{c}_c$) separately [6]. Compared to NH-CMF, V-CF more accurately describes pattern of palm elements: filter response's magnitude $|\vec{c}_f(x, y)|$ forms a finer pattern of extracted lines that closely corresponds to lines (veins / creases) in the input image. These filters are based on NH-CMF. Each correlation value c_{f_n} is calculated as in (4) using scaled kernel that matches palm's biometric feature, and additional non-linear artifact removal unit is used to remove responses that originate from processing line endings, thinner lines, edges and noisy areas. Thus, vein filter, which processes R channel image f_R extracts mainly (for 70.6% of the responses [6]) vein patterns, whereas crease filter, which processes B channel image f_B crease pattern (for 64.7% of the responses [6]). For the rest of calculated responses, filters detect features of opposite modality, other palm lines or noise. The percentage is calculated taking area and response magnitudes into account [6], which both are important for HOV.

4.2. Extraction of HOV features

The literature on feature extraction is very vast, but in general feature extraction trends can be divided into two classes: learning-based (LE [3], SFRD [4]) and hand-crafted features (LBP [1], HOG [5]). Learning-based descriptors lead to the state-of-art performance in many recognition tasks, but the computational simplicity of hand-crafted approach is more desirable in our case. This choice is dictated by limited computational power of the proposed embedded biometric system. None of the above-mentioned
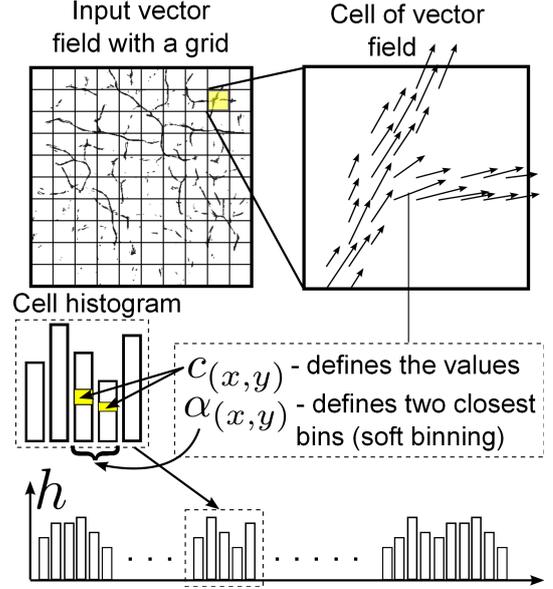

Figure 6. Extraction of HOV features

hand-crafted features can be directly applied to our recognition problem. The reason for that is the specificity of the input, which is a vector field obtained in the primary stage of the feature extraction process. Each pixel (x, y) in the input signal is described with two parameters: magnitude $c_{(x,y)}$ and angle $\alpha_{(x,y)} \in [0, 2\pi]$. In order to utilize both parameters a novel descriptor HOV is developed.

HOV approach inherits some ideas both from LBP and HOG descriptors. In order to save the spatial information about the object the vector field / image is divided into K^2 cells with a grid of the size $K \times K$ ($K = 10$ in Figure 6). Next, a spatially enhanced histogram $\boldsymbol{h}$ is calculated by sequential concatenation of the cell histograms into a single vector (similar to LBP approach [1]). The calculation of cell histogram is similar to HOG approach [5]. Angular values of each pixel define the corresponding bins/positions in the 1D cell histogram, while magnitudes define the amount to be added to the corresponding bins, Figure 6. The soft binning principle [5, 15] is incorporated in the process of cell histogram calculation. With soft binning each pixel benefits to more than one bin, assigning proportionally higher fractions of the magnitude to the closest bin and respectively smaller fraction to the distant bin.

The length of the HOV feature vector is $N = K^2 \cdot N_{\text{cell}}$, where N_{cell} is the number of bins in the cell histogram. The final step is normalization of the sum of histogram $\boldsymbol{h}$ elements to one: $\hat{h}_i = h_i / \sum_{j=1}^N (h_j), i = 1, \ldots, N$, where $\hat{h}_i$ are the entries of the normalized HOV histogram.
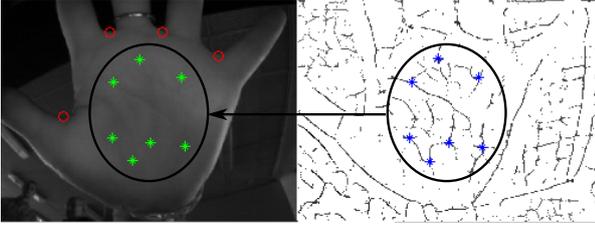
Figure 7. Example of ground-truth data for the database

5. Experimental results

Since the capturing of images requires special conditions (specific image sensor and illumination), the database of palm images captured with our developed device is introduced. The database contains data for $N_p = 100$ individuals with $N_{im} = 20$ images for each person resulting in 2000 images in total. Moreover, each palm image in the database is supplemented with ground-truth data including $P_{gap} = 4$ manually marked finger-gaps ('o'-markers in Figure 7) and $P_{ROI} = 7$ characteristic points on the palm ('*'-markers in Figure 7, hereinafter referred to as ROI points). The ROI points are linked to palm vein intersection, because these are easier to mark by the human experts. ROI points are in the same positions for all images of the person. Obviously, ROI points differ for different individuals, because of variability in the vein patterns. The database and ground-truth data are utilized in the evaluation of all algorithmic blocks of the proposed system. The database is publicly available for the research community upon request.

The detection, alignment and recognition stages of the algorithm and their mutual influence are evaluated next to understand the impact of each stage on the final recognition precision.

The result of the detection is represented by the binary mask covering the ROI in the input palm image, Figure 1, M. The database includes only those images, where proposed palm detection algorithm can properly detect the palm.

The general purpose of the alignment stage is to minimize the variance in appearance of the object of interest in images. Alignment is used to exclude mismatches resulting from rotation, translation and scaling. Obviously, in well aligned images the distribution of ROI points from their mean positions is minimal and vice versa. Ideally after alignment of 20 images of the person the corresponding ROI points should match perfectly. However, the ground-truth data incorporates some nonlinear deformations, which can not be corrected with affine transformations. Therefore, the inherent error of ground-truth data is analyzed first. For this purpose the ROI points of each person are aligned using RANSAC algorithm, Figure 8. Then the distances between aligned ROI points and their corresponding mean positions are calculated (distances $d_1, d_2, d_3, \ldots$ in Figure 8). These
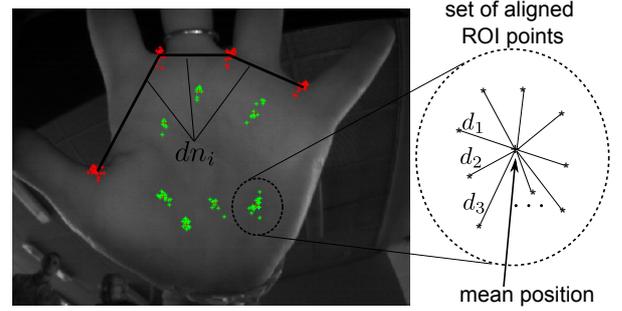

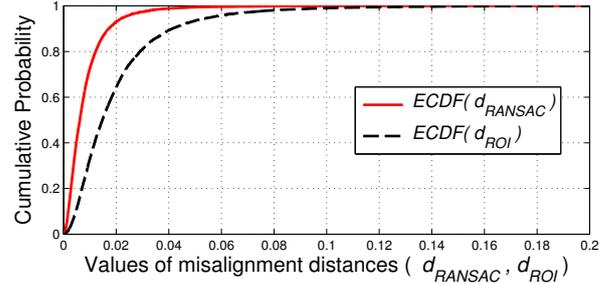Figure 8. Example of aligned ROI points using RANSAC

Figure 9. ECDF for $\boldsymbol{d}_{\mathrm{RANSAC}}$ and $\boldsymbol{d}_{\mathrm{ROI}}$ vectors

distances are next normalized and concatenated into a single vector $\boldsymbol{d}_{\mathrm{RANSAC}} = (\ldots, d_1/dn_1, d_2/dn_2, d_3/dn_3, \ldots)$, where the normalizer dn_i is the total distance between finger-gaps in the current image i, Figure 8. The length of the vector $\boldsymbol{d}_{\mathrm{RANSAC}}$ is $N_p \cdot N_{im} \cdot P_{\mathrm{ROI}} = 14000$, if the normalized distances are calculated for all ROI points in all images of the database. Next, the Empirical Cumulative Distribution Function (ECDF) is calculated for the data in vector $\boldsymbol{d}_{\mathrm{RANSAC}}$. The same analysis is performed for the proposed ROI detection and alignment algorithm. In this case the vector of normalized distances is designated as $\boldsymbol{d}_{\mathrm{ROI}}$. ECDF of $\boldsymbol{d}_{\mathrm{ROI}}$ represents the precision of alignment of each person's palm images with the proposed algorithm.

Corresponding ECDF curves are displayed in Figure 9. In the case of perfect alignment of vein-networks the ECDF would be a step function at value 0. In Figure 9 one can see that the alignment of vein-networks based on manually marked ROI points outperforms the proposed automatic ROI alignment algorithm. Impact of the localization errors on the recognition performance is also determined.

The analysis of final recognition stage is performed with the introduced database. The Equal Error Rate (EER) is selected as the performance criteria. The comparison of HOV feature vectors is based on Nearest Neighbor Classifier (NNC) incorporating histogram intersection as distance function. In order to identify the amount of error introduced by the automatic ROI detection module two sequences of experiments are performed. In the first sequence all images

Table 1. EER values in % for various sequences of experiments and filters

	NH-CMF + HOV + NNC				V-CF + HOV + NNC			
	Manual alignment		Automatic alignment		Manual alignment		Automatic alignment	
	Veins	Creases	Veins	Creases	Veins	Creases	Veins	Creases
HOV parameters	$K = 12$ $N_{\text{cell}} = 18$	$K = 12$ $N_{\text{cell}} = 18$	$K = 6$ $N_{\text{cell}} = 15$	$K = 6$ $N_{\text{cell}} = 15$	$K = 12$ $N_{\text{cell}} = 18$	$K = 12$ $N_{\text{cell}} = 18$	$K = 6$ $N_{\text{cell}} = 15$	$K = 6$ $N_{\text{cell}} = 15$
EER, %	0.99	3.33	1.89	5.15	**0.79**	**1.98**	1.84	3.62
EER,% after fusion					**0.77**		1.82	

in the database are aligned with RANSAC algorithm based on manually marked ROI points (expected misalignment error is represented with $\text{ECDF}(\boldsymbol{d}_{\text{RANSAC}})$ in Figure 9). The images are then normalized (to the same scale, position and rotation), filtered and cropped as described in Figure 1. This sequence is called manual alignment for simplicity. In the second sequence of experiments the processing pipe-line is exactly the same as introduced in Figure 1, meaning that all stages detection alignment and recognition operate in automatic mode. The expected misalignment error in this case is represented with $\text{ECDF}(\boldsymbol{d}_{\text{ROI}})$ in Figure 9. Hereinafter this sequence is called automatic alignment. HOV based recognition algorithm has two parameters to be optimized: grid parameter K and number of bins in the cell histogram N_{cell}. The following values are evaluated: $K = \{3, 6, 12\}$ and $N_{\text{cell}} = \{9, 12, 15, 18\}$. The size of the vector field is both sequences of experiments is 384×384 pixels. Vector field itself is calculated with two approaches: NH-CMF and V-CF.

The experimental results in Table 1 show, that specially designed V-CF filter outperforms more general NH-CMF approach. Also the EER in the automatic alignment mode is higher then the one in the case of manual alignment. The reason for that is misalignment error, which is greater for the proposed ROI detection approach, Figure 9.

For comparative purposes the EER is also obtained for HOG [5] feature vectors in combination with NNC. The distance function of the NNC is histogram intersection. The HOG descriptor can be explicitly defined with the following parameters taking into account that the overlap of the blocks is fixed at 50% [5]: η is the size of the square cell in pixels; ς is the block regioning factor meaning that each block is divided into $\varsigma \times \varsigma$ cells; β is the number of orientation bins in the histogram of the cell. The following values of the parameters are evaluated in the experiments both in manual and automatic alignment modes: $\eta = \{16, 32, 64, 128\}$, $\varsigma = 2$, $\beta = \{6, 9, 12, 15, 18\}$. The results are introduced in Table 2 showing the significant transcendence of the proposed HOV descriptor over a well known HOG approach.

The final stage of the recognition is fusion. According to the classification in [13] the fusion is done in the similarity score level. In this case the output classifier is using

Table 2. Comparison of HOG and HOV descriptors

Alignment	Manual		Automatic	
	Veins	Creases	Veins	Creases
HOG parameters	$\eta = 16$ $\beta = 12$	$\eta = 16$ $\beta = 18$	$\eta = 16$ $\beta = 18$	$\eta = 128$ $\beta = 15$
HOG EER, %	5.03	5.34	7.21	7.07
HOV EER, %	0.79	1.98	1.84	3.62

similarity scores obtained from two modalities as new features. The output of the combining classifier (perceptron in this case) is a new similarity score. In other words the perceptron tends to predict the class as either "genuine" or "impostor" based on the input similarity scores from two modalities. The introduced database is not split into training and test sets, therefore in order to avoid over-fitting only a small fraction of data (5%) is used for training and all available data is used for testing. The results are introduced in Table 1 and some of the values are highlighted in bold to demonstrate the lowest EER values.

6. Conclusion

To the best of our knowledge the introduced bi-modal palm biometric system is the first one capable of capturing two biometric parameters, palm veins and palm creases, synchronously with a single image sensor. Because of the special bi-spectral illumination required for the synchronous biometric parameter acquisition effective and simple detection of ROI in the palm image is possible. Since the capturing of images requires special conditions, the paper introduces a new palm image database. The database is supplemented with ground-truth data simplifying the error analysis of all algorithmic blocks.

The purpose of the paper is to introduce the combination of algorithms for bi-modal palm biometric system and to analyze the error propagation in the automatic recognition pipeline in order to identify the blocks to be improved. The experimental part shows that palm and ROI detection principle is robust. The precision of the automatic alignment stage is below the performance of the manual align-

ment utilizing the ground-truth data collected by the human expert, Figure 9. The misalignment error in the automatic recognition pipeline leads to a decreased performance: EER = 1.82% versus EER = 0.77% in the case of manual alignment, Table 1. The important conclusion is that approximately $(1.82 - 0.77)/1.82 = 58\%$ of the final EER (1.82%) comes from imperfectnesses in the alignment stage, while $0.77/1.82 = 42\%$ of the EER is due to feature extraction and classification module.

The proposed recognition results can be considered as a baseline for the introduced database collected with the novel multi-modal palm biometric system.

Acknowledgment

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *ECCV (1)*, pages 469–481, 2004.

[2] W. Bu, Q. Zhao, X. Wu, Y. Tang, and K. Wang. A novel contactless multimodal biometric system based on multiple hand features. In *Hand-Based Biometrics (ICHB), 2011 International Conference on*, pages 1–6, Nov 2011.

[3] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In *CVPR*, pages 2707–2714, 2010.

[4] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In *CVPR*, pages 3554–3561, 2013.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR (1)*, pages 886–893, 2005.

[6] T. Eglitis, M. Pudzs, and M. Greitans. Bimodal palm biometric feature extraction using a single rgb image. In *Proceedings of the 13th International Conference of the Biometrics Special Interest Group, 2014*, pages 39–50, September 2014.

[7] D. Gaikwad and S. Narote. Multi-modal biometric system using palm print and palm vein features. In *India Conference (INDICON), 2013 Annual IEEE*, pages 1–5, Dec 2013.

[8] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):4–20, Jan 2004.

[9] G. Michael, T. Connie, and A. Jin. Design and implementation of a contactless palm print and palm vein sensor. In *Control Automation Robotics Vision (ICARCV), 2010 11th International Conference on*, pages 1268–1273, Dec 2010.

[10] C. Poon, D. Wong, and H. Shen. A new method in locating and segmenting palmprint into region-of-interest. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 533–536 Vol.4, Aug 2004.

[11] M. Pudzs, R. Fuksis, R. Ruskuls, T. Eglitis, A. Kadikis, and M. Greitans. Fpga based palmprint and palm vein biometric system. In *Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the*, pages 1–4, Sept 2013.

[12] M. Pudzs, M. Greitans, and R. Fuksis. Complex 2d matched filtering without halo artifacts. In *18th International Conference on Systems, Signals and Image Processing (IWSSIP), 2011*, pages 1–4, June 2011.

[13] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24:2115–2125, 2003.

[14] J. Ruiz-del Solar, R. Verschae, and M. Correa. Recognition of faces in unconstrained environments: A comparative study. *EURASIP Journal on Advances in Signal Processing*, 2009(1):184617, 2009.

[15] F. Tang, S. H. Lim, and N. L. Chang. An improved local feature descriptor via soft binning. In *Proceedings of the International Conference on Image Processing, ICIP 2010, September 26-29, Hong Kong, China*, pages 861–864. IEEE, 2010.

[16] J. G. Wang, W. Y. Yau, and A. Suwandy. Feature-level fusion of palmprint and palm vein for person identification based on a "junction point" representation. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 253–256, Oct 2008.

[17] D. Zhang, W. K. Kong, J. You, and M. Wong. Online palmprint identification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1041–1050, Sept 2003.