Robust GrayScale Distribution Estimation for Contactless Palmprint Recognition

Julien Doublet, Marinette Revenu and Olivier Lepetit

Abstract—More and more research have been developed very recently for automatic hand recognition. This paper proposes a new method for contactless hand authentication in complex images. Our system uses skin color and hand shape information for an accurate hand detection process. Then, the palm is extracted and characterized by a robust and normalized decomposition. During enrollment, a distribution estimation is used to defined the optimal discrimination of the palmprint features. Finally, some specific thresholds are defined to separate in test phase impostor and genuine users. The experimental results present an error rate of 1.5% with a population of 49 people.

I. INTRODUCTION

Recently, biometrics recognition have attracted a lot of research and many works are reported in the literature. Biometrics technology is used to identify person from their physical or behavioral characteristics. These characteristics have three major specificities, to define each people, to be stable in the course of time and to be defined without a lot of constraints for users. So, various techniques were based on fingerprints, iris, face, voice, signature, gait or hand. Hand biometrics present many advantages compared to other biometrics technologies. Its characteristics are relatively stable and present a lot of discriminating features such as principal lines, wrinkles, ridges or hand shape. In addition, they present an high user acceptability and the hand can be obtained from low-resolution images with cheaper devices. Many algorithms have been defined for hand authentication based on palmprint [1], [2], the hand-shape recognition [4] or a combination of the two data [6], [5]. All these algorithms can be decomposed in four stages: the acquisition, the hand segmentation (and palm definition if necessary), the features extraction and the comparison with the reference. Generally, the acquisition is a hand scan. Then, in palmprint recognition, the hand is segmented by a fixed threshold and the palm is generally defined from hand shape curvature [2]. Next, various methods have been proposed for recognition. They are Sobel and morphological operations, Fourier spectrum, 2D Gabor phase [2], wavelet signatures [3], eigenpalm and eigenfingers features [7], phase and orientation code [8], etc. These papers present really good results in recognition rate using generally a comparison method based on Hamming distance adjusted following the features. In all these systems, the hand acquisition is in fact the principal limit. It is non-hygienic, all users touch the same glass, and some artifacts can be created during the acquisition according to the pressure of users on the plate glass.

In this paper, we propose a contactless recognition with a webcam to reduce these limitations and to get a quick and easy image acquisition without an additional device. Hand detection process from these acquisitions is a complex task. Many methods are defined for hand detection in man-machine interactions like skin blob tracking, active contours, mean shift or condensation algorithm but provide an approximate detection of each finger and often used tracking information. To increase the segmentation quality without tracking, we propose a method combining information of skin color and hand shape. The segmentation phase is described in section 2. In section 3, we propose a feature extraction by a robust decomposition and a palmprint matching using a distribution estimation. Section 4 contains some experimental results and section 5 provides the conclusions.

II. PREPROCESSING

Palmprint recognition preprocessing is used to define the central part of the hand (the palm). The most part of these preprocessing algorithms are decomposed in three steps. Firstly the hand is binarized by thresholding. Next, the hand contour is extracted and some key points are defined. Finally, from these points a coordinate system is defined and the palm is extracted. In our system, the architecture is kept but the methods are adapted for a contactless system.

A. Hand detection process

Preprocessing is used to define the central part of the hand (the palm). These algorithms are decomposed in three steps. Firstly the hand is binarized by thresholding. Next, the contour of the hand is extracted and some key points are defined. Finally, from these points a coordinate system is defined and the palm is extracted. In our system, the architecture is kept but the methods are changed for a contactless system.

1) Skin color segmentation : Classically the skin color is modeled by Bayes classifiers or Gaussian Mixture Models [9]. Contrary to these methods, our skin color is modeled by machine learning and for a good compromise between the execution time and the precision of detection, we use a neural network (NN). The NN entries are composed of three neurons, one for each color component of pixels in RGB domain. The NN output is the probability that a pixel is a skin pixel. The learning phase permits to model the skin
Many methods are defined using the probabilities map to clearly identify the detected skin object (hand or face) and to get an accurate segmentation. The principal method consists in segmentation by histogram thresholding on the probabilities map [10]. But more complex processes are defined in using watershed algorithm, median filtering or skin pixels grouping. In our approach, a morphological dilatation with a circular element is carried out on the probabilities map to group skin pixels. Next, the probabilities map is segmented by histogram thresholding. This threshold is fixed to 0.5: each pixel is considered as skin color if the neural network output value is more that 0.5 else it is classified as background tone. Finally, as our hand detection process will be used for biometrics recognition, we suppose that only one hand is presented to the system. So the region with the largest surface is preserved and considered as the hand. The regions with an important size and elongation are so conserved. They can be some fingers disconnected from the hand for the users wearing some rings. To reconnect these regions to hand, each region is approximated by an ellipse and we search in the direction of the major axis if the hand is close. If the region is near to the hand, an adjusted patch is added to reconnect the finger to the hand. This patch has a width equals to the width of the minor axis and a length depending on the distance between the finger and the hand. It is oriented with the major axis direction and centered between the hand and the finger. This process is illustrated in Fig. 1.

2) Color active shape model : The segmentation by skin color can not carry out the task of hand detection in an accurate way. So, a specific active shape model is defined to cancel this problem and to solve the two major difficulties of active shape model [11]. These problems are the contour initialization which must be close to the real form and the model convergence in the detection phase. Classically, the form to detect is defined by a set of points:

\[ X[0], X[1], \ldots, X[(M + 1) \times 10] \]

where \( X[i] \) is the ith landmark. After the initialization phase, the model is deformed. To control the problem of model divergence which does not follow the real hand contours, a weight is applied to deformations to limit the shape constraints. In addition, the gradient is computed in the skin color space by Di Zenzo algorithm [12]. Next, this gradient is balanced by the coefficient of the probabilities map pixels. So the background objects do not perturb the contour evolution. The experiments show that a good compromise between the execution time and the detection precision is obtained by fixing \( M \) at 12. This complete detection process is illustrated on Fig. 2(a-d).

B. Palm definition

After hand detection, it is necessary to extract the acquired palm independently of the distance between the hand and the capture device. Our extraction is based on hand dimensions and the palm extraction method described in [2]. In [2], two values are fixed: the distance between the points \( O1 \) and \( O2 \) and the palm size \( ||A1A2|| \) (Fig. 3). These constant values in traditional recognition systems are defined here according to hand sizes. They are determined from the hand width, computed by the Euclidean distance between the points \( X[L1] \) and \( X[L2] \) where \( L1 \) and \( L2 \) are the indexes

![Fig. 1. Fingers re-connection (a) binarized map (b) patch adjustment (c) binarized map with patch](image)

![Fig. 2. Hand detection (a) original image (b) skin segmentation (c) initial hand shape (d) final hand shape](image)
fixed after experiments at 30th and 125th landmarks. Thus, 
\[ |O1O2| \text{ and } |A1A2| \text{ are defined by:} \]

\[
\begin{align*}
|O1O2| &= \alpha |X[L1,X[L2]|  \\
|A1A2| &= \beta |X[L1,X[L2]| 
\end{align*}
\]

(1)

Where \( \alpha \) and \( \beta \) are the dimension coefficients chosen at 1/10 and 2/3, respectively. Then the palm is resized with a fixed size \( N \times N \).

The extracted palm contains some principal lines which can be determined by a specific palm line extraction. These lines are not unique to each individual, thus it is necessary to use the ridges and the secondary lines of the palm. This information cannot be extracted from the palm with images in low resolution, so a global palm characterization is suitable.

III. ROBUST FEATURES EXTRACTION

From preprocessing, the palm features must be extracted. A lot of method have been proposed to define the specific information of each user. The most important methods consist of statistical approach (decomposition into small regions and computation of the means and the variance of each region), subspace-based approach (Principal Component Analysis, Independent Component Analysis or Discrete Cosine Transform) and filtering approach (encoding filter responses). In all these systems the image noise is generally not taken into consideration as the acquisition is performed in a constraint environment and with a specific device.

In our system the sensor is limited (webcam or camera phone). So, we proposed to remove the noise by a PDE (Partial Differential equations) regularization [13]. Moreover in the majority of statistical systems and subspace-based approaches, the luminosity changes are managed by a global histogram equalization or a normalization of the palm to a fixed average and variance. As these methods are not robust to inhomogeneous luminosity changes, we proposed a normalization by a palm decomposition. These two methods are described in the following paragraphs.

A. Palm restoration

The images acquired with low quality devices present important noises (typically following a Poisson distribution) and compression artifacts specially in an environment with difficult lighting conditions. These two types of perturbations are difficult to remove and many methods tends to resolve this problem. A well tool to make a powerful regularization of noisy images is the anisotropic diffusion PDE’s. PDE regularizations are based on locale image smoothing. This smoothing is depending on the gradient information: with important gradient, the smoothing is effecting along the gradient direction else, in homogeneous region, the smoothing is performed in all directions. So, these methods give an anisotropic regularization without edges destruction.

Recently, Tschumperlé [13] proposed an anisotropic smoothing of images using curvature-preserving PDE’s fast and maintaining the thin image details. This method is based on a separation of the smoothing process respecting the local smoothing geometry. Moreover, the author defines a regularization given into consideration the contours curvature. It’s important to note that this regularization is applied only in palm restoration and not for the total image restoration to limit the computing time. In fact, the multi-resolution computation in skin detection process and the hand shape learning limit the noise impact in the hand detection process.

B. Palm decomposition and normalization

From the restored palm, a local normalization is made to be robust of inhomogeneous luminosity change. Let \( I_l \) an image \( I \) with an additional uncontrolled luminosity impact and \( I(x,y) \) the gray level at point \( (x,y) \). \( I_l(x,y) \) can be defined by:

\[ I_l(x,y) = I(x,y) + L(x,y) \]  

(2)

where \( L(x,y) \) defined the luminosity impact at the point \( (x,y) \). An approximation perhaps carried out by taking into account the fact that close points have a similar luminosity. So, the definition (2) could be defined by:

\[ I_l(x,y) = I(x,y) + L_r(x,y) \]

(3)

where \( L_r(x,y) \) is the luminosity impact in a region near the point \( (x,y) \). Using this definition, the palm image is normalized by a decomposition into adjacent small regions that are supposed to have the same luminosity impact. Next, these regions are standardized.

Firstly, a palm image is decomposed into \( \Lambda \) regions defining a set of sub-images \( \{R_0,R_1,\ldots,R_\Lambda \} \). The number of regions depends on the sub-image size \( S \times S \) and the intersection width between each region \( W \).

Next, each region \( R_i \) is normalized to a specified mean and variance. The mean \( \bar{\phi}_i \) and the variance \( \rho_i \) of each region \( R_i \)
are computed before standardization. With these data, each normalized region $R_i$ is determined by:

$$R_i = \begin{cases} 
\phi_i^d + \lambda(x, y) & \text{if } R_i(x, y) > \phi_i \\
\phi_i^d - \lambda(x, y) & \text{otherwise}
\end{cases} \quad (4)$$

where

$$\lambda(x, y) = \sqrt{\frac{\rho_i^d (R_i(x, y) - \phi_i)^2}{\rho_i}}$$

and $\phi_i^d$ and $\rho_i^d$ are the desired mean and variance of the region $R_i$, respectively. So, the palmprint feature $R'$ is a set of normalized sub-image $\{R'_0, R'_1, \ldots, R'_\Lambda\}$.

As each user has different palms, the desired data (mean and variance) are defined individually for each user to keep the most possible information. In a learning phase, these specific data are fixed by the median value of variance and mean of palmprint regions for each user. Figure 4 shows an illustration of decomposition and normalization process applied to a restored palm with the parameters $S = N/2$ and $W = N/4$.

IV. REFERENCE DEFINITION BY DISTRIBUTION ESTIMATION

From the palmprint features, a decision function must be defined to verify if two references are similar. We propose to use a classification method. With its good accuracy, we used an estimation of the support of a high-distribution estimation [14] based on the Support Vector Machine (SVM) introduced by Vapnik [15].

A. Distribution estimation

Initially, the SVM permits to classify some elements in different classes with linear functions by projecting the data in a higher dimensional space. In [14], Scholkopf and al. proposed an extension of the support vector algorithm in the case of unlabeled data in estimating the support of a high-dimensional distribution.

Given a training data $\{x_1, \ldots, x_l\}, i = 1, \ldots, l$ where $l$ is the number of examples, $x_i \in \mathbb{R}^m$ is the example of index $i$ and $m \in \mathbb{N}$ is the data dimension. Let $\Phi$ a function of projection $\mathbb{R}^m \rightarrow F$, a kernel function $K(x_i, x_j) = (\Phi(x_i), \Phi(x_j))$ is defined. An example of the kernel function is the radial basis function defined by:

$$K(x_i, x_j) = \exp^{-\gamma ||x_i - x_j||^2}, \gamma > 0 \quad (5)$$

The goal of the method proposed by Scholkopf and al. is to determine a function $f$ that return $+1$ in a region near the training data and $-1$ elsewhere. The objective is to map the data in a feature space and to separate them with maximum margin. The data set can be separated from the origin in solving the following primal problem:

$$\min_{w, \xi, \rho} \frac{1}{2} w^T w + \frac{1}{l} \sum_{i=1}^{l} \xi_i - \rho$$

subject to $w^T \Phi(x_i) \geq \rho - \xi_i, \xi_i \geq 0. \quad (6)$

where $\xi \in \mathbb{R}^l, \rho \in l, w \in F, v \in [0, 1] \text{ controls the number of support vectors and the errors.} v$ is an upper bound on the fraction of the outliers (training points outside the defined region) and a lower bound of the fraction of support vectors. The dual problem of (6) can be obtained in introducing a Lagrangian. This dual can be explained by:

$$\min_{\alpha} \frac{1}{2} \sum_{ij} \alpha_i \alpha_j K(x_i, x_j)$$

subject to $0 \leq \alpha_i \leq \frac{1}{vl} \sum_{i=1}^{l} \alpha_i = 1. \quad (7)$

where $\alpha_i \geq 0$. Using the Lagrangian, the decision function $f$ can be explained with the kernel function by:

$$f(x) = \text{sign} \left( \sum_{i=1}^{l} \alpha_i K(x_i, x) - \rho \right) \quad (8)$$

where $\text{sign}(a) = \begin{cases} +1 & \text{if } a \geq t \\
-1 & \text{otherwise}
\end{cases}$
Classically, the threshold \( t \) determining the decision is fixed to the value 0.

### B. Individual Authentication function

In our palmprint matching, we propose to use a decision function for each user. So, in learning phase a decision function \( f_{I_d} \) is estimated for each client \( I_d \). The palmprint features of each user \( R_{I_d} \) defined in section III are transformed into a 1D vector \( X_{I_d} \) by concatenation each element of the reference \( R_{I_d} \), row by row. \( \Gamma \) references \( X_{I_d}^j, j = 1, \cdots, \Gamma \) defined of class +1 are used as training data to defined this estimation. Because the palm extraction process is imperfect, each palmprint feature used to define the decision function is translated (vertically and horizontally) and rotated to supplement the training data. The ranges of the translation and the rotation are defined from \(-2^\circ\) to \(2^\circ\) respectively.

The learning phase is fast with so few examples but it’s not possible to determine a same fixed thresholding \( t \) for all users to separate impostor and genuine because the decision function is only depending on a set of positive examples. So, for each user \( I_d \) a threshold \( t_{I_d} \) must be defined. This threshold \( t_{I_d} \) is computed in projecting the references \( X_{I_d}^j, j = 1, \cdots, \Gamma \) into the function \( f_{I_d} \). Next, the median value of these projections is kept and adjusted with a value \( \tau \) to define the threshold. This definition can be explained by:

\[
  t_{I_d} = \text{median}_j(f_{I_d}(X_{I_d}^j)) - \tau
\]

The parameter \( \tau \) limits the impact of bad palm extraction and noisy images in matching phase. The median value is used to get the notion of variation between a genuine palmprint feature and the decision function and to be robust to a training feature obtained with an important bad palm acquisition.

### V. EXPERIMENTATIONS

Two databases are defined to validate our approach. The first one permits to test the hand detection process and the second one the recognition system.

#### A. Hand detection evaluation

The database to skin detection is composed of 2189796 palm skin color pixels and 3599970 non-skin color pixels in different illumination conditions. The hand skin color pixels coming from different ethnic groups are used to train our neural network in different color space. The better results are obtained in RGB domain with a classification error rate equals to 6.2% in our test database of 1215600 pixels equally distributed among skin and non-skin pixels. This good classification can be explained by the fact that the palm skin color is relatively stable for all ethnic groups and the neural network can define a non-linear modeling.

The model used in active shape model is learned from 86 segmented and marked hand images by saving 98% of information contained in the first 13 eigenvectors. An another hand images set composed of 87 marked images enable to fix the optimal ASM values. The complete experimentations of hand detection process (without fingers connection method) are presented and discussed in [16].

#### B. Palmprint recognition evaluation

A specific database is created to validate the recognition approach. All images of the database were acquired by a webcam Philips ToUcam Pro 740K with a size of 640 × 480 and automatic color adjustment to limit the impact of illumination change. The database contains 490 images of hands, some of these presenting hands with rings, coming from 49 people. 10 images are acquired in one session for each individual of the database. The capture restrictions are that the users must present their hands in front of the camera with theirs fingers separated. The users are not accustomed to the system, so in practice, a bad acquisition is automatically rejected by the system. The palm is extracted by our hand detection process and resized to \(128 \times 128\) \((N = 128)\). If the initial palm size is lower than \(128 \times 128\), the image is rejected to ensure a minimal image quality. Figure 5 shows two examples of hand images with corresponding palms.

Let \( X_{I_d}^j \) the palmprint feature of the palm \( j \) of the user \( I_d \). First, each user \( I_d \) is described by a function decision \( f_{I_d} \). The training data for each user to fix this function and the variances and means which permit the normalization of the decomposed palm are defined by the set \( \{X_{I_d}^j\} (j = 1, \cdots, \Gamma) \) of class +1. In evaluation phase, a test database is defined for each user. This database is explained for an identity \( I_d \) by a genuine set \( \{X_{I_d}^j\} (j = \Gamma + 1, \cdots, 10) \) of class +1 and an impostor set \( \{X_{I_d}^j\} (e = 1, \cdots, 49, e \neq d, j = 1, \cdots, 10) \) of class −1. All these data are standardized by the means and the variances defined in learning phase. To increased the number of test, the complete process of experimentations (learning and evaluation) is done ten times in randomly changed the index \( j \) of each reference.

In the complete evaluation phase, the rates of classification are determined. A comparison is incorrect if a user is accepted or rejected wrongly. Two rates characterize the
accuracy of the system: the false acceptance rate (FAR) defined by the ratio of people authenticated wrongly and the false rejection rate (FRR) defined by the ratio of people rejected wrongly. All recognition rates presented in this paper correspond to 100 – equal error rate (EER) defined when \( \text{FAR} = \text{FRR} \).

Different values for the system parameters \( \Gamma, K \) and \( \tau \) are used to obtain the best recognition rate. The parameter \( \Gamma \) explains the number of examples to form training set. When the threshold adjusted by \( \tau \) varies the recognition rates are different. Table I shows the system performance following these two data. It is found that the recognition rate is higher than 1 second on Pentium M with 1.60GHz. The enrollment time is depending on the number of training examples. With \( \Gamma = 5 \), the decision function computation is done in 4.7 seconds.

VI. CONCLUSIONS

In this paper, we present a new method of biometrics hand recognition for a contactless system. First of all, the hand segmentation is explained. It is carried out thanks to an integration of the skin color components and a shape model. Then, the authentication process based on robust features extraction and a specific distribution estimation is exposed. The features extraction permits to get an information robust to illumination change and limited noise. Next, an individual threshold is defined for each specific decision function. The complete process is validated after experiments on a database of 49 people. It shows an error rate of 1.5% with an execution time lower than 1 second.

In order to manage space rotations of the hand, an invariant method with the perspective and the shear will have to be required. Moreover, a definition of the value \( \tau \) for each user must be defined. These two tracks are currently the subject of our work.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>( \Gamma = 3 )</th>
<th>( \Gamma = 4 )</th>
<th>( \Gamma = 5 )</th>
<th>( \Gamma = 6 )</th>
<th>( \Gamma = 7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>81.5%</td>
<td>83.5%</td>
<td>84.7%</td>
<td>85.8%</td>
<td>86.1%</td>
</tr>
<tr>
<td>0.03</td>
<td>85.2%</td>
<td>88.1%</td>
<td>92.5%</td>
<td>93.4%</td>
<td>93.6%</td>
</tr>
<tr>
<td>0.04</td>
<td>92.5%</td>
<td>94.9%</td>
<td>96.2%</td>
<td>97.0%</td>
<td>97.3%</td>
</tr>
<tr>
<td>0.05</td>
<td>93.7%</td>
<td>96.0%</td>
<td>97.3%</td>
<td>98.4%</td>
<td>98.5%</td>
</tr>
<tr>
<td>0.06</td>
<td>88.1%</td>
<td>92.5%</td>
<td>92.7%</td>
<td>93.3%</td>
<td>93.6%</td>
</tr>
<tr>
<td>0.07</td>
<td>83.2%</td>
<td>84.2%</td>
<td>85.9%</td>
<td>86.3%</td>
<td>86.4%</td>
</tr>
</tbody>
</table>

Table I: Compare the performances of the system following the number of training samples and the adjustment threshold \( \tau \).

Fig. 6. Recognition rates with different kernels.