International Journal of Security and Its Applications
Vol. 9, No. 5 (2015), pp. 335-346
http://dx.doi.org/10.14257/ijsia.2015.9.5.33

Complex Background Palm Segmentation using SVM

Gao Peng[1] and Li Liang[2]

1School of Foreign Language, Xidian University, Xi'an, China
2Xi'an Research Center, Huawei Technologies Co., Ltd, Xi'an, China
pgao@xidian.edu.cn, ralf.liliang@huawei.com

Abstract

An efficient method is proposed to enhance complex background segmentation for pre-processing in palmprint recognition. In this paper, we integrate texture and haze features by applying Laws masks to the image represented in YCbCr color space, and use these features of a patch with its neighborhood information to determine hand and non-hand regions by support vector machine (SVM). Compared with other methods, our algorithm demonstrated robustness to changing illumination and a complex environment and we obtain a relatively clear hand shape with an average accuracy of 94.96%. The images in our experiments are taken with popular mobile phones in our laboratory.

Keywords: *Background segmentation, Palmprint recognition, Laws masks, Support Vector Machines (SVM)*

1. Introduction

Palmprint recognition, which has been widely applied in the areas of network security and identity authentication [18], usually utilizes CCD-based palmprint scanners as the input sensors. As users have to put their hands on the sensors to acquire hand images, latent palm prints remaining on the surface are vulnerable to copying for illegitimate use [9]. Without touching the sensor, contactless devices are more flexible and hygienic and, more importantly, the contactless method is effective against the illegal use of palm prints.

Compared with traditional security authentication approaches, such as keys or passwords, biometrics are more accurate, dependable and more difficult to steal or fake. A typical palmprint recognition system consists of four parts: palmprint collector, preprocessing, feature extraction and matcher. Hand segmentation will directly impact on the accuracy of the region of interest (ROI) cutting, and is related to feature extraction and matching. With environment changes, the difficulty of segmentation is enhanced and the application of this biometric technology is limited. The existing recognition systems have a good performance when the background is simple or constant but are less effective in a complex environment. It is a challenging task to segment the hand from a variable background.

In this paper, we give an innovative background segmentation technique based on using the texture energy and haze of a patch with its neighborhood information, which is applied to contactless palmprint recognition in complex background conditions. Firstly, the YCbCr color space is utilized because of its robustness to illumination variation. We choose features that capture two types of local cues: texture variations, and haze using Laws masks [4]. Also, we include the features of four neighbors of a patch to increase determinacy in classification. This is discussed in detail in section 3. In the classification phase, due to its excellent performance on dealing with classification problems, SVM is used for training and making predictions. We define two classes for our problem: hand region and non-hand region labeled with +1 and -1 separately.

ISSN: 1738-9984 IJSIA

The rest of this paper is organized as follows: Section 2 lists the existing background segmentation methods; Section 3 presents the features we have extracted; Section 4 gives a brief introduction to SVM; Section 5 evaluates and compares the performance of our method and Section 6 concludes the whole paper.

2. Related Work

In early years, palmprint recognition systems captured hand images via the ink method or palm scanners. Limited by the simple acquisition mechanisms, the background is single or fixed as in [7]. The fixed threshold method is efficient for binarization in such cases and is a cost effective solution.

To cope with a more complex environment, Otsu [10] proposed an adaptive threshold selection method based on the histogram of a picture. The main idea of the approach is applied to the gray level picture and based on the assumption that the object and the background in a single image follow a normal distribution. The algorithm can choose the valley bottom between two peaks representing objects and background as the threshold of an image to maximize the differences between foreground and background. Zhu, *et al.*, [20] used this method while Poon, *et al.*, [13] employed an adaptive threshold technique, using the statistical information of the background and the hand regions, to segment the hand image from the background.

With palmprint recognition systems being widely used, traditional segmentation technologies cannot meet the needs. The photos we take usually contain different colors and different objects in the background. To deal with complex environments, researchers focused their attention on skin color detection. It is observed that the features of the hands cluster in RGB color space and other color spaces such as Hue-Saturation-Value (HSV), Normalized RGB (NRGB), YUV and CIELAB. Many methods have been proposed to maximize the separability of skin and non-skin regions. In 2006, Doublet, *et al.*, [5] segmented the hand from the background using RGB color space. In that paper, an artificial neural network (ANN) is used to determine the threshold in RGB space. Yoruk, *et al.*, [19] used the K-means clustering algorithm on both the gray-level pixels and the RGB components to separate the hand foreground and the darker background. However, RGB is sensitive to illumination variation so that the algorithm is not robust in a changeable environment. Rotinwa-Akinbile, *et al.*, [14] proposed a novel palmprint segmentation technique using Cb and Cr channels of skin and non-skin pixels. The ANN is used to train the 2-input neural network architecture to create a model, and the K-means algorithm and coefficients in the neural network model is used to predict which part of a test image is hand.

Some research on background removal adopted other features of a hand image. For example, in Han, *et al.*, [6], the Haar features are extracted to solve the two-class problem of classifying the input photos into hand and non-hand classes. In that system, with a training set of left-hand and non-left-hand shapes, including right hand, human head, cars and turtles, etc., Han employed the Float AdaBoost algorithm to select the most efficient Haar features to construct the cascade classifier to solve the problem.

However, the methods mentioned above are not suitable for images with variations in illumination and environment. For example, the adaptive threshold method has a reduced performance when the background contains different types of colors and objects; the results of skin color detection are not fine enough when illumination varies; and the method in Han, *et al.*, [6] is limited by the specific equipment required. Hence, in this paper, we propose a novel background

segmentation solution, based on texture energy, haze and neighborhood information, which can effectively deal with various environments and illumination changes.

3. Features

Unlike the methods mentioned, our aim is to deal with palmprint recognition on a mobile phone with a built-in camera, where hand placement has a high degree of flexibility. In this paper, we will use the results of filtering by Laws masks on YCbCr color space and its neighborhood information as features to classify hand and non-hand regions.

3.1. Feature Acquisition

In order to cope with the variations in illumination and environment, we integrate texture and haze features. Texture information is typically contained within the image intensity channel and haze is reflected in the low frequency information in the color channels. Since YCbCr color space is known for its ability to separate luminance and chrominance, we extract these features by applying Laws masks to these three channels.

Laws masks [4] are usually used to generate texture features which can detect various types of textures. Here we use the 3×3 convolution masks which are derived from one-dimensional vectors of three pixels length. L_3, E_3, S_3 describing the level, edge and spot features respectively are as follows:

$$\begin{cases} L_3 = [1 \quad 2 \quad 1] \\ E_3 = [-1 \quad 0 \quad 1] \\ S_3 = [-1 \quad 2 \quad -1] \end{cases} \tag{1}$$

The nine Laws' masks are generated as:

$$\begin{matrix} L_3L_3 & L_3E_3 & L_3S_3 \\ E_3L_3 & E_3E_3 & E_3S_3 \\ S_3L_3 & S_3E_3 & S_3S_3 \end{matrix} \tag{2}$$

Nine different templates can be generated by above 1×3 vectors (Figure 1).

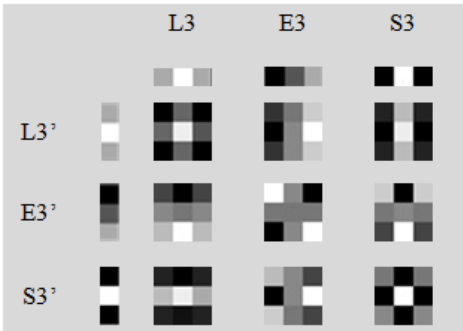


Figure 1. Nine Laws' Masks

We apply these masks to obtain filtered results representing texture energy of a patch. All of the masks are used for intensity channel I_Y and for color channels I_{Cb} and I_{Cr}, we employ the first Laws mask only [8]. Then we can gain a total of 11 texture features of a patch [8, 15-17]. These filtered images can be obtained by

$$\begin{cases} I_n = I_Y * M_n, \ n = 1, 2, \ldots, 9 \\ I_{10} = I_{Cb} * M_1 \\ I_{11} = I_{Cr} * M_1 \end{cases} \tag{3}$$

where M_n is the nth template.

3.2. Neighborhood Structure

We divide the image into small patches, and predict the classification of a patch. We denote the feature vector of patch i by

$$F_i = \sum_{(x,y) \in patch(i)} I(x, y) \tag{4}$$

where $I(x, y)$ is the output of the eleven filters.

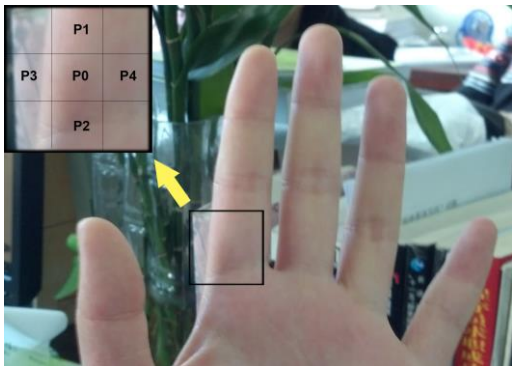


Figure 2. One Patch P0 and its Neighboring Patches

It is observed that information in each patch of an image cannot exist independently. The class a specific patch belongs to is closely related to its neighborhoods (Figure 2). Whether a patch in an image is hand or not can be predicted by the patches near it. For example, if all of its neighborhoods belong to foreground, we can believe the patch is also foreground. In the case where there are more neighbors included in background, it is more likely that the patch belongs to background. In Figure 2, most of the neighboring patches of P0 belong to the hand region, so we have increased confidence in stating that P0 is hand. Since the color features of human skin cluster in YCbCr space and are robust to illumination, we can estimate the separability between two patches easily. Therefore, we use color features of one patch and of its four neighboring patches as features.

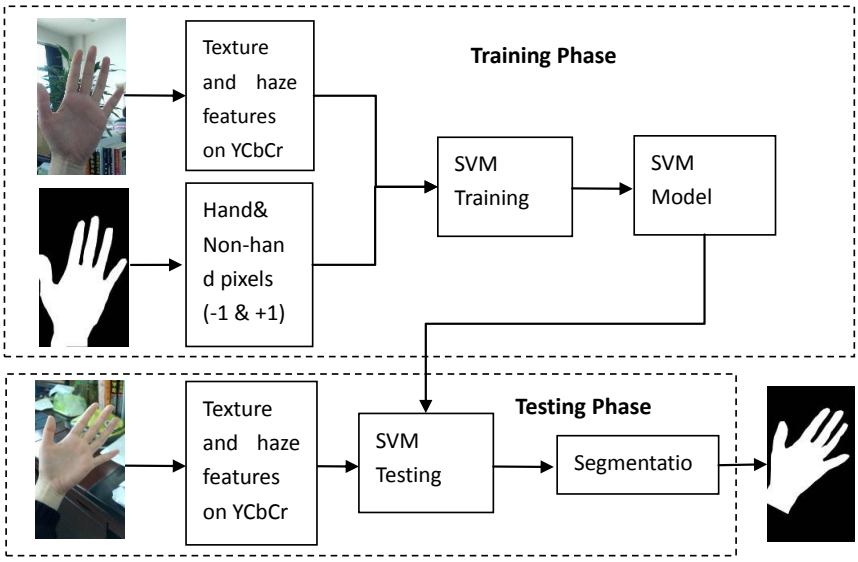


Figure 3. Flow Diagram of the Proposed Technique

4. Classification: Support Vector Machines

Due to its excellent performance on classification, ability to manage high dimensional space problems and avoidance of local optimal solutions, SVM has been used in a large range of classification problems, such as face detection [12], and image classification [2]. In this section, we provide a brief review of the theory behind this type of algorithm. For more details, we refer the readers to [3].

The input of SVM is a training set $(x_i, y_i)_{1\le i\le m}$, where $x_i \in R^n$, n being the dimension of problem space, is features of an example and correspondingly, yi is the target labeled by +1 or −1. Our aim is to find a hyper plane, also called decision surface, which can divide the training examples into two parts and also correctly predict an unknown example to which class it belongs. The hyper plane can be represented as

$$w^T x + b = 0 \tag{5}$$

w is a vector that is perpendicular to the decision surface. In order to obtain an optimum, we need to maximize the minimum distance between the hyperplane and the points in the training set. So the target function is

$$\max_{1\le i\le m} \frac{y_i(w\cdot x_i + b)}{\| w \|} \tag{6}$$

$\| w \|$ is the Euclidean norm of w. To simplify the problem, according to the Lagrange Duality and the Kuhn Tucker condition, we can convert the target function to a simple dual quadratic optimization problem:

$$\begin{cases} \max \quad W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ s.t. \quad \sum_{i}^{m} \alpha_i y_i = 0, \quad \alpha_i \ge 0, i = 1,2,...,m \end{cases} \tag{7}$$

where $a_i \ge 0$ is the Lagrange multiplier and $\sum_{i=1}^{m} a_i y_i = 0$.

When it comes to a linear non-separable case, SVM can be extended by using kernel $K(x_i, x_j) = <\Phi(x_i), \Phi(x_j)>$ that satisfied Mercer condition. The purpose of introducing kernel is to map the input space into a higher dimension space so that we can find a decision surface to classify examples, as in the linear separable case. The objective function is

$$\begin{cases} \max \quad W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \quad \alpha_i \ge 0, i = 1,2,...,n \end{cases} \tag{8}$$

In our work[1], a RBF kernel $K(x, y) = \exp\left\{ \frac{-\| x - y \|^2}{2\sigma^2} \right\}$ was used, and we applied an SVM network to our experiment (Figure3). Each patch in every image here is regarded as one training example. In our learning algorithm, the input is the features we have extracted and the target $y_i \in \{+1, -1\}$ labeled manually, representing either a hand region or non-hand region. Due to the different types of

[1] During the feature analysis phase (section 5.1), we use the Scikit-learn for SVM in python [11]. In the performance evaluation phase (section 5.2), we use the LIBSVM [1].

backgrounds in our dataset, we managed to guarantee images belonging to each type were trained. Also, in order to avoid any existing bias caused by several final images that are in the same type, training examples are input in a random sequence.

5. Experiment Result and Evaluation

For the evaluation, we collected 120 images from 7 individuals hands with various illumination and random backgrounds for our experiments. We used the MOTO ME865 to acquire our dataset. The device is equipped with an 8MP camera that takes 1840*3264 pictures with 72dpi. We planned for 70% of the images to be used for training and the remaining 30% for testing.

In this paper, the performance of the method was evaluated using:

$$Accuracy = \frac{TP + TN}{P + N} \tag{9}$$

where *TP* is the number of patches correctly classified as hand region, and *TN* is the number of patches correctly classified as non-hand region. P is the total number of hand patches while *N* is the total number of non-hand patches.

5.1. Feature Analysis

During the feature analysis phase, we use 16 samples in the dataset for training in order to make a trade-off between accuracy and efficiency. Figure4 shows the absolute value of correlation coefficients between eleven feature values and targets. The 1st, 10th and 11th features are more significantly correlated with the target than others. The second to the ninth features have higher correlation with each other (Table 1) while they are insignificantly related to the target. Owing to the redundancy between these eight features, there's no additional gain in adopting all of them. Thus, the best three features (1st, 10th and 11th) should be employed, and we select some of the others to provide supplementary benefits to gain a better performance.

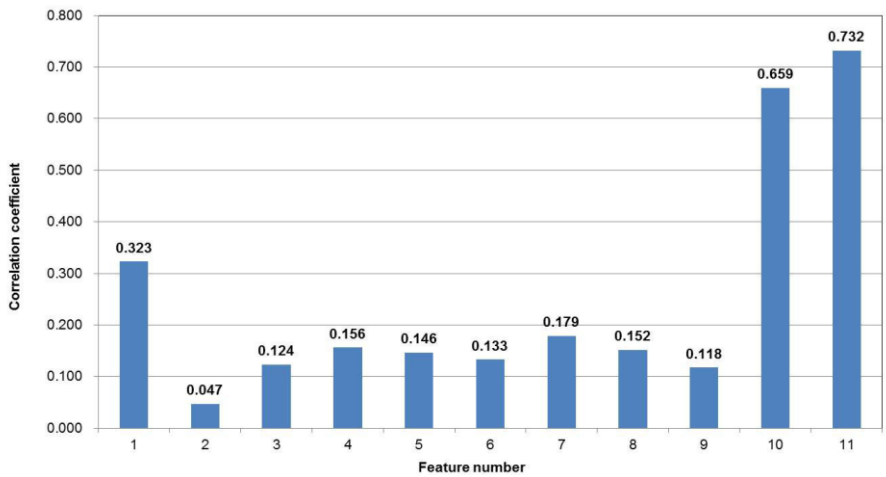


Figure 4. The Absolute Value of Correlation Coefficients between Feature Values and Targets

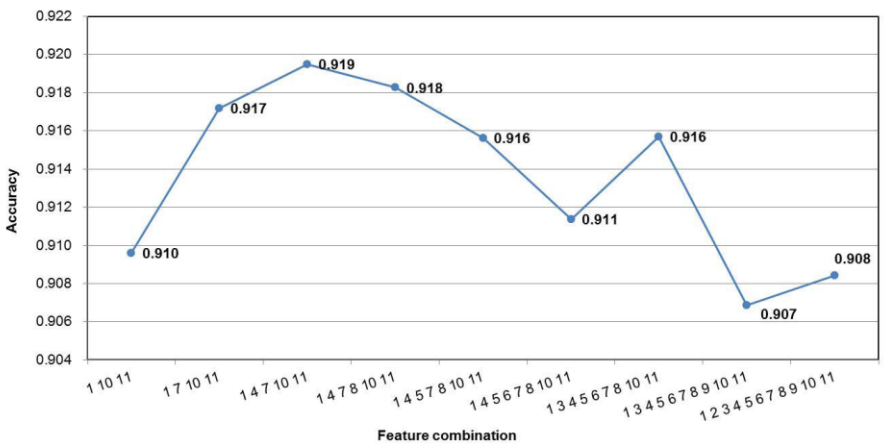


Figure 5. Accuracies of Different Feature Combinations. Every Time the Added Feature has the Maximum Correlation Coefficient among the Rest According to Figure 4

Table I. Correlation Coefficients between Different Feature Values

	1	2	3	4	5	6	7	8	9	10	11
1	1.00										
2	0.00	1.00									
3	-0.02	0.70	1.00								
4	-0.02	0.33	0.21	1.00							
5	-0.03	0.38	0.41	0.45	1.00						
6	-0.02	0.50	0.44	0.39	0.50	1.00					
7	-0.03	0.18	0.18	0.59	0.42	0.30	1.00				
8	-0.02	0.29	0.26	0.48	0.50	0.41	0.46	1.00			
9	-0.02	0.33	0.36	0.35	0.44	0.53	0.39	0.52	1.00		
10	-0.07	0.02	0.07	0.13	0.12	0.11	0.15	0.13	0.10	1.00	
11	-0.04	-0.01	-0.07	-0.13	-0.12	-0.10	-0.15	-0.13	-0.10	-0.84	1.00

According to the correlation coefficients between feature values and targets (Figure 4), we have done a number of feature combination experiments. Figure 5 presents several combinations where the number of features sequentially increases. Each time the remaining feature with the maximum correlation coefficient is added. Note that the combination of the best three features (1,10 and 11) and the 4th and 7th features achieve the highest scores.

Hence we chose the combination of features 1, 4, 7, 10 and 11, and give the comparison of experiments using these features with and without neighborhood information. It is shown that the introduction of neighborhoods can enhance the performance and precision of segmentation (Figure 6), which gives a clearer outline of the palm. The average accuracy in the two cases is 91.87% and 90.37%, respectively. Neighborhoods features performed a little better than non-

neighborhoods features. In particular, features with neighbors could segment the area between fingers more accurately.

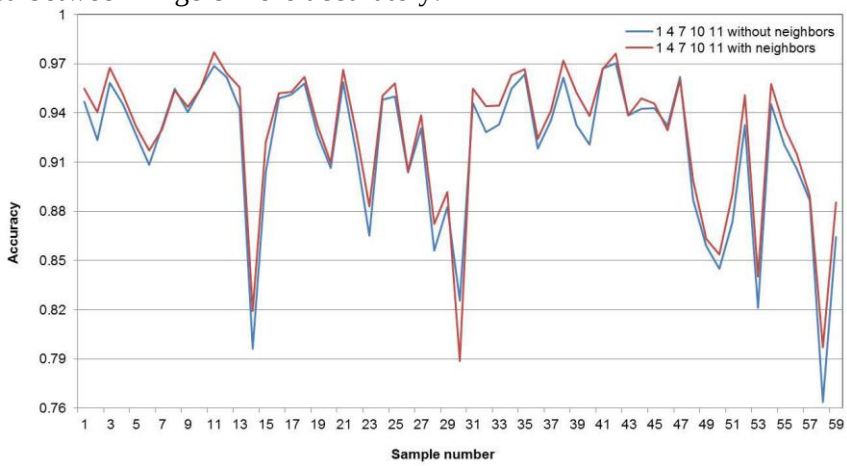
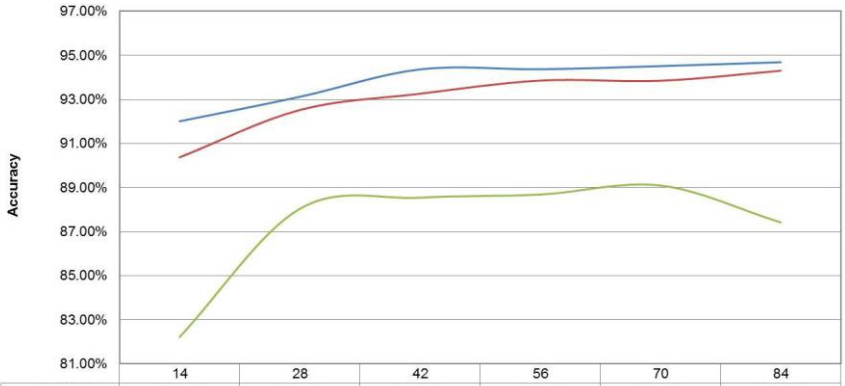


Figure 6. Accuracy of Features with and without Neighbors

5.2. Performance Evaluation

In order to appraise the performance of this technique, the algorithms in [14, 5] were also executed for comparison. We use (9) to evaluate the performance of these methods. The results of our method and the algorithm in [14, 5] for different sizes of training sets are shown in Figure 7. It illustrates that even in the case of a small number of training samples, our method is still performing well. The segmentation results for the algorithms are shown in Figure 8. The algorithm presented in this paper has a good qualitative performance compared with the other two methods. Seen in Figure 8, the performance of the palmprint recognition system has been greatly enhanced that edge lines of the palm are smoother and clearer than the other two methods.



	14	28	42	56	70	84
Our Method	92.01%	93.12%	94.37%	94.37%	94.51%	94.69%
Doublet	90.37%	92.52%	93.26%	93.86%	93.85%	94.30%
M.O.Rotinwa	82.22%	88.04%	88.54%	88.69%	89.09%	87.42%

Figure 7. Accuracies of the Three Methods for Different Training Sample Size

6. Conclusion and Recommendation

In this paper, we have proposed a novel method of hand segmentation. Our goal is to extract a more complete hand shape from variable backgrounds and overcome environment restrictions. It is based on a combination of the texture energy feature, haze, and neighborhood features. The palm characteristics are computed by Laws masks under intensity and color channels. For classification, an SVM based on the target vectors for

the given feature dataset is employed. The complete process is validated via our experiments, demonstrating an accuracy level of at least 92%. Our algorithm demonstrates a better performance than other techniques and the method is well-tested using a variety of images with random background.

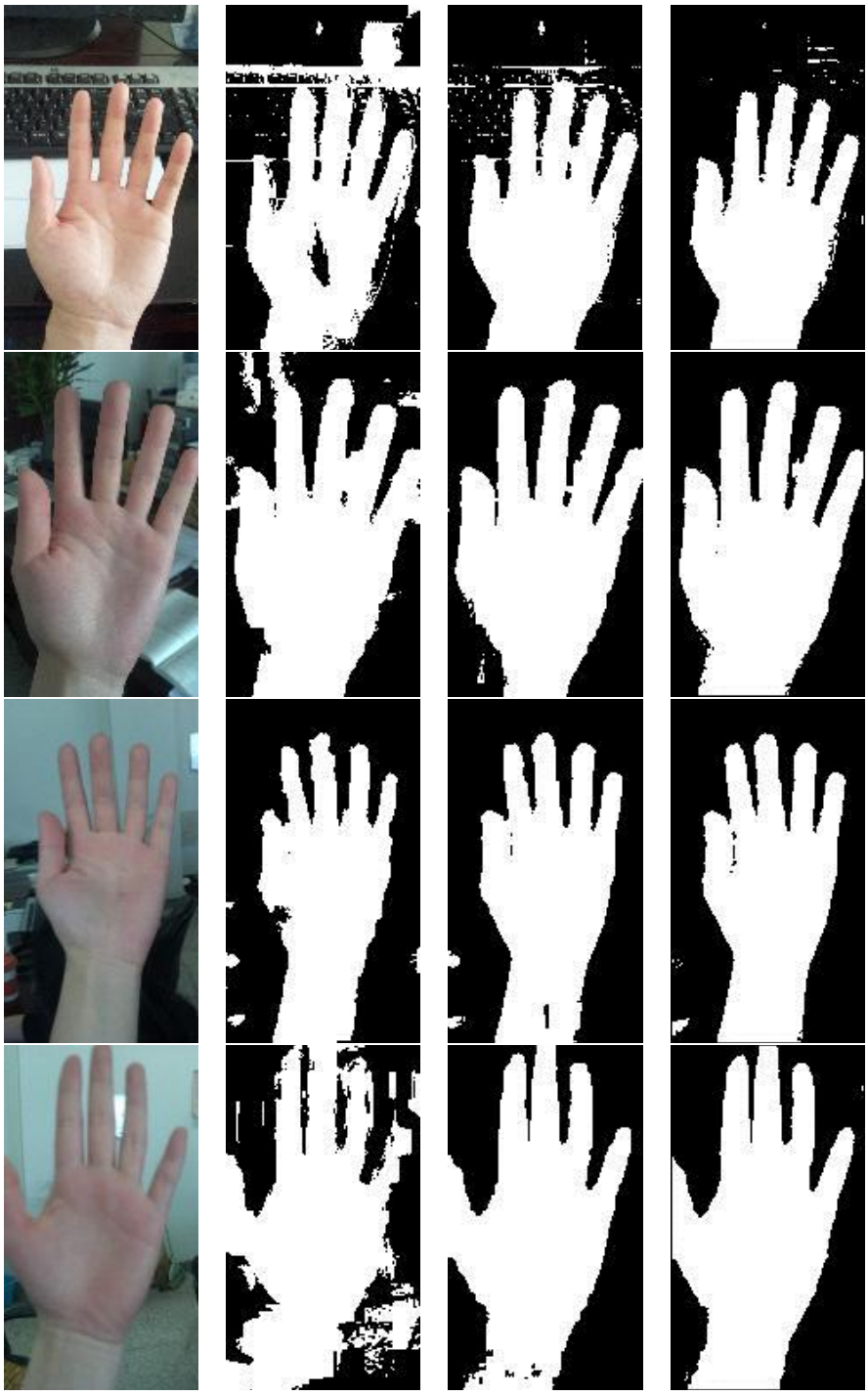


Figure 8. Results of Segmentation Algorithms, showing Original Image (Column 1), Predicted Image by BP Neural Network and K-means Algorithm with CbCr Features [10] (Column 2), Predicted Image by BP Neural Network with RGB Features [8] (Column 3), and Predicted Image by our Method (Column 4)

ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for their careful reading of this paper and for their helpful and constructive comments. This project is supported by the xxx.

References

[1] C. C. Chang and C. J. Lin, "Libsvm: a library for support vector machines", ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 27, **(2011)**.
[2] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification", Neural Networks, IEEE Transactions on, vol. 10, **(1999)**, pp. 1055–1064.
[3] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods", Cambridge university press, **(2000)**.
[4] E. R. Davies, "Machine Vision: Theory", Algorithms, Practicalities and Edition. Acad emic Press, San Diego. chapter Laws texture energy in TEXTURE, **(1997)**.
[5] J. Doublet, O. Lepetit and M. Revenu, "Contact less hand recognition using shape and texture features", in: Signal Processing, 2006 8th International Conference on, IEEE, **(2006)**.
[6] Y. Han, Z. Sun, F. Wang and T. Tan, "Palmprint recognition under unconstrained scenes", in Computer Vision–ACCV, Springer, **(2007)**, pp. 1–11.
[7] W. Li, D. Zhang and Z. Xu, "Palmprint identification by fourier transform", International Journal of Pattern Recognition and Artificial Intelligence, vol. 16, **(2002)**, pp. 417–432.
[8] J. Michels, A. Saxena and A. Y. Ng, "High speed obstacle avoidance using monocular vision and reinforcement learning", in: Proceedings of the 22nd international conference on Machine learning, ACM, **(2005)**, pp. 593–600.
[9] M. G. K. Ong, T. Connie, A. B. J. Teoh, "Touch-less palm print biometrics: Novel design and implementation", Image and Vision Computing, vol. 26, **(2008)**, pp. 1551–1560.
[10] N. Otsu, "A threshold selection method from gray-level histograms", Automatica, vol. 11, **(1975)**, pp. 23–27.
[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss and V. Dubourg, "Scikit-learn: Machine learning in python", The Journal of Machine Learning Research, vol. 12, **(2011)**, pp. 2825–2830.
[12] P. J. Phillips, "Support vector machines applied to face recognition", Citeseer, **(1998)**.
[13] C. Poon, D. C. M. Wong and H. C. Shen, "A new method in locating and segmenting palmprint into region-of-interest", in Proceedings of the 17th International Conference on Pattern Recognition, IEEE, **(2004)**, pp. 533–536.
[14] M. Rotinwa-Akinbile, A. M. Aibinu and M. J. E. Salami, "A novel palmprint segmentation technique", in: Informatics and Computational Intelligence (ICI), 2011 First International Conference on, IEEE, **(2011)**, pp. 235–239.
[15] A. Saxena, S. H. Chung and A. Ng, "Learning depth from single monocular images", Advances in Neural Information Processing Systems, vol. 18, no. 1161, **(2006)**.
[16] A. Saxena and A. Y. Ng, Make3d, **(2007)**, http://make3d.cs.cornell.edu/.
[17] A. Saxena, M. Sun and A. Y. Ng, "Learning 3-d scene structure from a single still image", in: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, IEEE, **(2007)**, pp. 1–8.
[18] J. Wayman, A. Jain, D. Maltoni and D. Maio, "Biometric systems: technology, design and performance evaluation", Springer, **(2005)**.
[19] E. Yoruk, E. Konukoglu, B. Sankur and J. Darbon, "Shape-based hand recognition", Image Processing, IEEE Transactions on, vol. 15, **(2006)**, pp. 1803–1815.
[20] X. Zhu, D. Liu and Q. Zhang, "Research of technology for thenar palmprint locating and segmentation", in: Proceedings of 2009 International Workshop on Information Security and Application, **(2009)**, pp. 601–604.

Authors

Gao Peng received her Bachelor's degree and Master Degree at 2005 and 2008 separately from Xi'an International Studies University, she is now a faculty lecturer of School of Foreign Language, Xidian University. Her research interests include linguistics, visualization and computer security.

Li Liang received his Bachelor's degree at 2005 from Xidian University, he is now a senior software engineer of Xi'an Research Center, Huawei technologies Co. Ltd. His research interests include wireless communication and computer security.

