DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-291

Contactless palm landmark detection and localization on mobile devices

Yaqi Wang, Liangrui Peng, Shengjin Wang, Xiaoqing Ding
Tsinghua National Laboratory for Information Science and Technology, Dept. of Electronic Engineering, Tsinghua University, Beijing, China

Abstract

Palmprint recognition as a novel biometric identification method for contactless mobile devices has been received substantial attentions in recent years. Palm landmark detection is one of the key technologies of palmprint identification and verification system. However, the differences of hand positions, complex backgrounds and various lighting conditions in unrestrained environment with low-resolution cameras make palm landmark detection in the wild difficult. In this paper, we proposed a new palm landmark detection approach based on Supervised Descent Method (SDM). SDM uses the relationship between the feature representation and the position of a landmark point to build an optimization problem for palm landmark detection. The optimization target function is the distance of feature representations between current position and the ideal position of a palm landmark point. After optimization, a linear function of the position displacement and the feature representation of current landmark is obtained. The linear function can be learned from palmprint image samples with labeled landmark positions. Given an input image in detection process, the initial position of a landmark is set by the mean position of the landmark in the training set, then the optimal landmark position can be calculated iteratively using the learned linear function. The effectiveness of the proposed method is proved on a mobile phone captured palm image dataset.

1. Introduction

In recent years, personal recognition on mobile devices is becoming increasingly important. Biometrics, one of the most reliable methods in this area, including face, voice and fingerprint, has been widely used. Compared with other biometric features, palmprint as a relatively new biometric recognition method contains more information and needs cheaper capture devices. Palm landmark detection plays a critical role in getting better rates and selection of region of interest for palm recognition and identification. Most approaches use databases collected from fixed image acquisition equipment with monotone or restricted backgrounds to get rid of the complex localization and segmentation problems [10]. Systems based on those approaches thus require usage of physical restraints to guarantee consistent hand positioning. However, the problem is challenging when hand images are taken by touch-less and unrestricted systems with extreme poses, lightings, and backgrounds, which becomes an obstacle of the popularity of contactless palmprint recognition system on mobile devices [8].

Existing approaches of palm location and segmentation in complex backgrounds can be generally divided into two parts: the pixel-based approaches and the model-based approaches. Color detection is a typical pixel based example that has been adopted in [1], [2]. The disadvantage of this method is that the images cannot be separated exactly when there are objects with skin color in the background. The model-based approaches fit a generative model for the global hand appearance. Doublet [3] used the active shape model (ASM) [4] for the contact-less palmprint system. The ASM approach chooses a set of shape parameters for a Point Distribution Model (PDM), calculates the main template of variance of the PDM and fits test image to the templates iteratively. However, ASM uses only shape information and its performance is thus impacted. Murat Aykut [6] developed another popular model based method for palm location called Active Appearance Model (AAM) [5]. AAM improves the accuracy of landmark localization of ASM by using a combined statistical model of shape and texture. The weakness is that AAM and its extensions are difficult to optimize.

In this paper, we employ Supervised Descent Method （SDM） to palm location in the palmprint recognition and identification system. SDM was first proposed by Xiong et al. [7] for minimizing a nonlinear least squares function basing on the Newton’ s method. As a typical optimization tool, Newton’s method plays an important role in smoothing twice-differentiable functions. Mathematical optimization algorithms such as Newton’s method can solve many detection and location problems in computer vision. However, Newton’s method cannot be directly applied to landmark detection problem for the following reasons: (1) The Newton steps will sometimes be taken in the wrong direction because the Hessian matrix can be positive somewhere in addition to the local minimum point. (2) Considering of the large dimension of the Hessian matrix, it will be computationally expensive when inverting or estimating the gradient of the Hessian matrix. For solving these problems, Supervised Descent Method is used to learn the descent directions in a supervised manner. During training, SDM learns a sequence of optimal descent directions. In testing, SDM minimizes the nonlinear least squares objective using the learned descent directions without the need for computing the Jacobian and Hessian. Experimental results show that SDM achieves great efficiency by tackling those two common troubles faced by 2nd order descent method mentioned above. With the location of landmark points on the palm determined from the fitted SDM, we can segment the region of interest and extract features for identification and verification. Moreover, our approach allows palm pictures taken in various environment.

The organization of this paper is as following. Section 2 introduces the SDM based method and discuss the algorithm used to fit a model to palm landmark detection. Section 3 describes the experimental results from the palm dataset that we collected under various backgrounds and lighting conditions. And the final conclusions are drawn in section 4.

2. Method

2.1. Mathematical derivation of SDM

Suppose we have an image $\mathbf{I} \in \Re^{M*N}$ of M*N pixels. There are p landmarks in the image. Each landmark has a 2-D coordinate $\boldsymbol{x}^{\boldsymbol{i}}$ (m^i, n^i) and all p landmarks are denoted by $\boldsymbol{x} \in \Re^{2p*1}$. $\emptyset(\boldsymbol{x})$ is

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-291

the nonlinear feature descriptor of the landmarks $\boldsymbol{x}$ such as SIFT features [9]. The SDM consists of two stages: training and testing. In the training part, location of the correct p landmarks $\boldsymbol{x}_*$ are known. The initial configuration of the landmarks $\boldsymbol{x_0}$ was represented by the average location of the training data. Landmark location can be converted to minimizing the objective function over offset $\Delta\boldsymbol{x}$:

$$f(x_0 + \Delta x) = \|\emptyset(x_0 + \Delta x) - \emptyset_*\|^2 \quad (1)$$

In this function, $\emptyset(x)$ represents the nonlinear feature values (SIFT) extracted from patches around the landmarks x. $\emptyset_* = \emptyset(\boldsymbol{x}_*)$ is the feature values of labeled landmarks in training data. There are some illustrations of Eq.1: (1) In the training images, the correct landmarks are labeled already. Based on the calculated average location of landmarks $\boldsymbol{x_0}$, $\emptyset_*$ and $\Delta\boldsymbol{x}$ are both known. (2) Instead of learning any model appearance in advance from training data, Eq.1 just need to calculate average landmark locations $\boldsymbol{x_0}$ and optimize the landmark locations $\boldsymbol{x}$ directly. This non-parametric shape model can generalize better to fit multi-pose palm situations. (3) The nonlinear SIFT operator we use is not differentiable. Using first or second order methods to minimize Eq.1 will lead to numerical approximations of Jacobian and Hessians which are computationally expensive. Therefore SDM learns a series of descent directions and rescaling factors to update the landmark positions $(\boldsymbol{x_{k+1}} = \boldsymbol{x_k} + \Delta\boldsymbol{x_k})$, to make the positions calculated from training data converge from $\boldsymbol{x_0}$ to the correct position $\boldsymbol{x}_*$ step by step.

Assume that $\emptyset(\boldsymbol{x})$ is twice differentiable for derivation purpose. Then a second order Taylor expansion was applied to Eq.1 as follow:

$$f(x_0 + \Delta x) \approx f(x_0) + J_f(x_0)^T \Delta x + \frac{1}{2}\Delta x^T H(x_0)\Delta x \quad (2)$$

where $\mathbf{J_f(x_0)}$ and $\boldsymbol{H(x_0)}$ refer to the Jacobian and Hessian matrices of function $\boldsymbol{f}$ at $\boldsymbol{x_0}$.

Take the derivative of both sides with respect to $\Delta\boldsymbol{x}$ and set the derivative at $\boldsymbol{x_0} + \Delta\boldsymbol{x}$ to zero to minimize the function $\boldsymbol{f(x_0} + \Delta\boldsymbol{x})$, then we can get an update of x:

$$J_{f(x_0+\Delta x)} = J_{f(x_0)} + H(x_0)\Delta x \quad (3)$$

$$\Delta x_1 = -H^{-1}J_f = -2H^{-1}J_\emptyset^T(\emptyset_0 - \emptyset_*) \quad (4)$$

We omit $\boldsymbol{x_0}$ in Eq.4 and the following equations to simplify the derivation. And $\boldsymbol{J_f} = \boldsymbol{J_\emptyset^T}(\boldsymbol{\emptyset_0} - \emptyset_*)$ is obtained from the relationship between $\boldsymbol{f}$ and $\emptyset$ shows in Eq.1 using chain rule, where $\boldsymbol{\emptyset_0} = \emptyset(\boldsymbol{x_0})$. So the update value $\Delta\boldsymbol{x}$ can be explained as the projection of $\emptyset_* = \emptyset(\boldsymbol{x}_*)$ onto $\boldsymbol{R_0} = \boldsymbol{-2H^{-1}J_\emptyset^T}$ in another way. It shows the linear relationship between $\Delta\boldsymbol{x}$ and the difference of the feature values $\Delta\emptyset$.

Therefore, by learning $\boldsymbol{R_0}$ from training data, the calculation of Jacobian and Hessian matrices can be avoided. So $\boldsymbol{R_0}$ can be regarded as a descent direction. In addition, the function $\boldsymbol{f}$ is not limited as twice differentiable. During testing, $\emptyset_*$ of the real landmark is unknown but fixed. So we can convert Eq.4 to the following form:

$$\Delta x_1 = R_0\emptyset_0 + b_0 \quad (5)$$

In training part, SDM will learn the parameters $\boldsymbol{R_0}$ and $\boldsymbol{b_0}$ in the above-mentioned update procedure. The details of learning process will be explained in the next section.

In general, the function $\boldsymbol{f(x)}$ is more complicated than quadratic polynomials, so during training, we will get several offsets $\Delta\boldsymbol{x_k}$:

$$\Delta x_{\mathrm{k}} = R_{\mathrm{k-1}}\emptyset_{\mathrm{k-1}} + b_{\mathrm{k-1}} \quad (6)$$

where $\boldsymbol{\emptyset_{k-1}} = \emptyset(\boldsymbol{x_{k-1}})$ is the SIFT feature values at the landmark locations getting from (k-1)$^{\text{th}}$ step. $\boldsymbol{x_k}$ will converge to the exact landmark location $\boldsymbol{x}_*$ iteratively:

$$x_k = x_{k-1} + R_{k-1}\emptyset_{k-1} + b_{k-1} \quad (7)$$

2.2. Hand landmarks detection based on SDM

2.2.1 Preprocess

According to the theory analysis above-mentioned, SDM uses linear model to build the relationship between coordinates and feature space. But the simple form will easily lead to over fitting. The training set should be large enough to reduce the susceptibility of initial configuration. However, the training data we collected is not that big. So we need to add disturbance to the training samples to reduce the susceptibility.

Add disturbance to each image from the training data $\{\boldsymbol{I^i}\}$:

$$I_k^i = A(a_k, u_k, v_k)I^i = \begin{bmatrix} a_k & 0 & u_k \\ 0 & a_k & v_k \end{bmatrix} I^i, k = 1,2 \ldots n \quad (8)$$

where $\boldsymbol{A(a_k, u_k, v_k)}$ is the perturbation matrix.

2.2.2 Feature extraction

The next step is feature extraction from key points. SIFT descriptors are chosen because it is invariant to uniform scaling, orientation, and partially invariant to affine distortion and illumination changes. In order to achieve orientation invariance, the coordinates and gradient orientations are rotated relative to the key point orientation. During experiment, it is found that the most stable result is obtained when the SIFT descriptors are computed on 16*16(4*4 histograms) local patches around each key point. Each histogram has 8 bins covering the 360 degree range of orientations. Therefore, each SIFT descriptor is a vector with 4*4*8=128 dimensions.

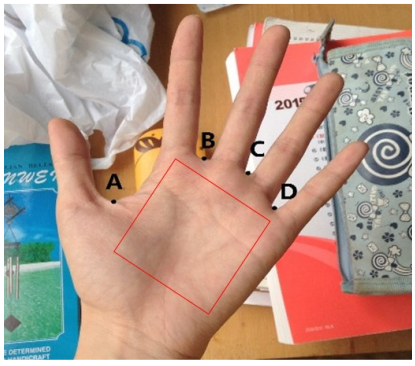


Figure 1. Representation of the region of interest (red box).

DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-291

Figure 2. Examples of landmarked images from the palm dataset. Hand positions and poses vary significantly between users.

2.2.3 Training phase

The core step of SDM training is to learn the descent direction $\boldsymbol{R}$ and constant $\boldsymbol{b}$. Given the training image set $\{\boldsymbol{I^i}\}$ and their corresponding hand-labeled landmarks $\{\boldsymbol{x_k^i}\}$, $\boldsymbol{x_0^i}$ is the initial position given by centering the mean position at the normalized square of all the images from training set. Solve $\boldsymbol{R}$ and $\boldsymbol{b}$ by minimizing the L2-loss function:

$$\underset{R_k b_k}{\operatorname{argmin}} \sum_{I^i} \sum_{x_k^i} \left\| \Delta x_*^{ki} - R_k \emptyset_k^i - b_k \right\|^2 \tag{9}$$

Applying the update rule in Eq.7 with $\boldsymbol{R_{k-1}}$, $\boldsymbol{b_{k-1}}$ learned from the last step, new $\Delta \boldsymbol{x_*^{ki}}$ and $\emptyset_{\boldsymbol{k}}^{\boldsymbol{i}}$ will be computed, which means we generate a new training data. At the first step, we initialize k=0 and $\Delta \boldsymbol{x_*^{0i}} = \boldsymbol{x_*^i} - \boldsymbol{x_0^i}$.

As the number of iteration increase, the error between the current position and the hand-labeled landmark will decrease. The experiment showed that the results tend to converge after about 5 steps.

In practical training, the premature convergence cases occur because of limited data size, which will influence the length of $\{\boldsymbol{R_k}\}$ and $\{\boldsymbol{b_k}\}$ and worse the detecting results.

2.2.4 Detecting phase

During detection, given the test image $\boldsymbol{I_t}$, descent direction $\{\boldsymbol{R_k}\}$ and constant $\{\boldsymbol{b_k}\}$ learned from training set and mean position $\boldsymbol{x_0^t}$, the aim is to calculate the landmarks according to the model trained from training set and the information from test images. The specific approach is as follow: Calculate the feature vector extracted at present landmark locations from $\boldsymbol{x_0^t}$, iteratively update the current position using Eq.7, and then output the convergent result.

2.2.5 ROI extraction

Since the purpose of the palm landmarks detection is correctly selected the palmprint region, which called Region of Interest (ROI) as shown in Figure 1. We choose 4 points at the valleys between the fingers (Point **A, B, C, D**). These 4 points are also chosen by many other approaches because they are sufficient for ROI determining. Line up Point ***B*** and Point ***D*** to get the X-axis of the ROI coordinate system. Draw a line passing through Point ***C*** which is perpendicular to the X-axis, and another line passing through Point ***A*** which is parallel to the X-axis. The cross point of those two lines is the center of the ROI square. The length of the side is set according to the length of Segment ***BD***. After normalization, the extracted ROI images can be sent to a palmprint recognition system.

3. Experimental Results

3.1. Hand image dataset

We set up a dataset[1] of hand images in complex background for development and testing of the palm graph based landmarks detection approach. 1,112 images were collected from 75 individuals. Images were taken by smartphone cameras with a fixed size of 512*512. Examples of the images are shown in Figure 2. In order to facilitate the performance of the proposed method realistically, palm of each person was imaged with different scenarios including the change of backgrounds, positions, rotation degrees and lighting conditions. More specific requirements are as follows: (1) People are asked to present their palms toward camera with spread fingers. The whole hand from fingertips to the heel of the hand should be contained in the image. (2) Each person is asked to put his or her left hand in different complicated backgrounds and white backgrounds. (3) Users are also asked to rotate their hands with different degrees and different distance to the camera. (4) To obtain the images in various lighting conditions and make the sample more representative, we captured the images under different occasions and in different time. After image collection, we marked the landmarks manually.

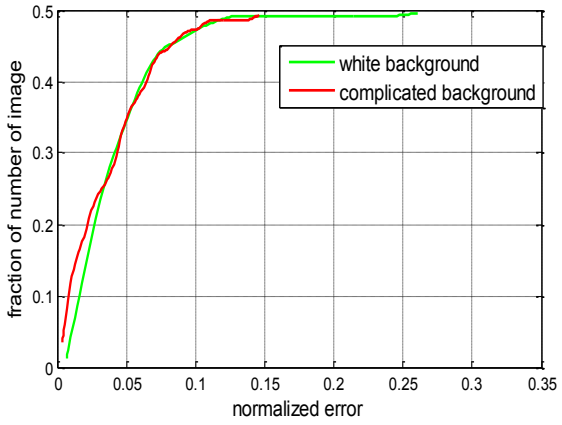
3.2. Detection accuracy

We use the point-to-point errors metric to measure the performance. For each image, normalize the Euclidean distance between the detected landmark points ($\boldsymbol{x^t}$) and corresponding reference landmark points($\boldsymbol{x_*^t}$) by the distance from the first valley point between the thumb and forefinger to the fourth valley point between the little and ring fingers ($\boldsymbol{d_{AD}}$). The relative error ($\tilde{\boldsymbol{d}}$) represents the accuracy of the algorithm:

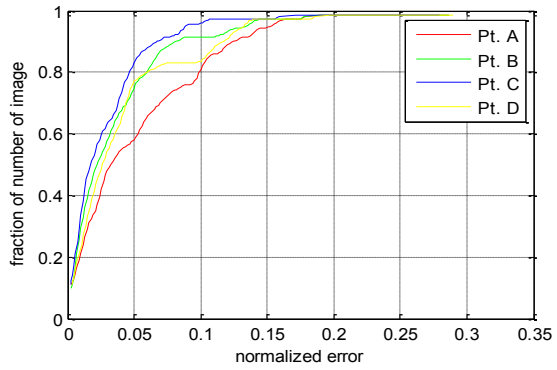
$$\tilde{d} = \left\| \frac{x^t - x_*^t}{d_{AD}} \right\| \tag{10}$$

[1] The dataset is available upon request for research purpose.

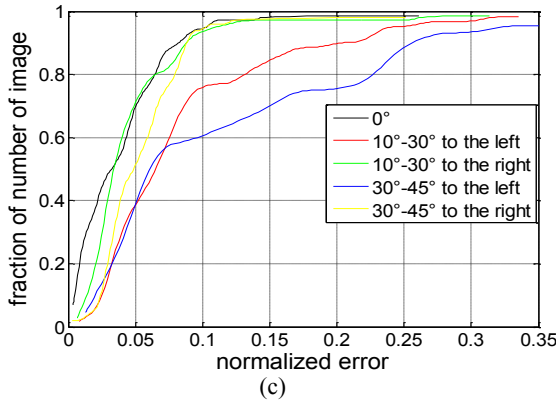
©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-291



(a)



(b)



(c)

Figure 3. Cumulative error curves on the test set. (a) Errors of the palms in white background and complex background are compared. (b) Errors of the 4 points are compared. Point A represents the landmark between the thumb and the forefinger. Point B and Point C represents the two valley points around the middle finger. The fourth valley point between the little and ring fingers is represented by Point D. (c) Errors of the palms with different degree of rotations are compared.

Table 1. Comparisons of average normalized errors.

Background	Complex	White
$\tilde{d}$	0.05	0.04

(a)

Point	Pt. A	Pt. B	Pt. C	Pt. D
$\tilde{d}$	0.08	0.04	0.04	0.05

(b)

Rotation degree	0°	10° - 30° (left)	30° - 45° (left)	10° - 30° (right)	30° - 45° (right)
$\tilde{d}$	0.04	0.05	0.06	0.09	0.13

(c)

After calculating the errors of all the images in the testing set, cumulative accuracy curves are given to show the percentage of images which have normalized errors under a threshold level in all the testing images. The reliability of detection ability increase with the steepness and rate of climb of the cumulative curves.

We conducted several experiments on the palm dataset we set up. The dataset was divided into two parts as training set and test set. Training set contains 356 images which were carefully selected to ensure the representation ability of the variations on the shape and background of the palm images. And the left 756 images are contained in the test set. Two performance metrics are used on the test set: One is the average normalized error mentioned above. The second one is the cumulative error curve. Figure 4 shows the visualized detecting results.

We classify the testing set in order to compare the performances in different conditions including the backgrounds and the degrees of hand rotation.

We compared the performances on the images taken in white background with images taken in complex background. Figure 3(a) shows the results on both two sets. It can be seen that the method can get good results in complex background as well as white background. We also compared the performance of different rotation degrees on the test set. The result can be seen at Figure 3(b). The rotation of the palms within a range of 30 degree has little influence of the detection accuracy. Figure 3(c) shows the validation error for the four different landmark points. The average normalized error comparisons are shown in Table 1. (Table 1. (a) Gives the normalized error values of different type of backgrounds. Table 1. (b) Gives the normalized error values of each points. Table 1. (c) Gives the normalized error values of with different degrees of rotations.)

DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-291

The proposed method was compared with the AAM method that utilizes shape and texture information [6] [11]. In [6], the maximum deviation from the landmark points of the reference model are 4.5%, while ours is 4.1%. It can be seen that the proposed method achieved higher accuracy than the AAM method. Furthermore, most of the methods like AAM method mentioned above are tested on images that collected from fixed camera systems with relatively fixed postures. While our dataset has the most complex background and the least constraints of postures, lighting conditions and rotations. It can be concluded that the proposed method has better performances and can be used in more complex situations. In terms of time complexity, the proposed method introduced a little overhead. The detecting stage is real-time. The algorithm is implemented in C++ and takes 0.05 second to process one image on an Intel i7-3700 CPU.

Figure 4. Some examples taken from the testing set. Our method is able to handle images that contain great variation in pose and background condition.

3.3. Error analysis

It can be noticed from Table 1 that different orientations of palms to the left and right have different performances. The reason is that because our system is applied on the smartphone and most users are right-handers, when taking the palm images by themselves, they tend to lean their left hands to the right. In order to adapt to the reality conditions, we put more right-slanting hands into the training set than the left-slanting hands.

The first point between the thumb and the forefinger has a worse detection accuracy compared with other points, as shown in Figure 5. It can be attributed to two reasons. The change of the arc between the thumb and the forefinger is larger than the other three points. In addition, due to the wide radian, the point-to-point error measurement is not very suitable because the landmark cannot be defined at a specific point.

However, the first point is not as important as the other 3 points in finding the ROI of palm, the worse performance of the first point will not lead to a bad influence for the follow-up work of palm recognition and identification.

Figure 5. Examples of inaccurate test results. The first point between the thumb and the forefinger has a worse detection accuracy.

4. Conclusions

This paper shows that SDM is suitable for palm landmark detection in unconstrained environment. The performance of the method is tested on the dataset, which has been constituted with the hand images taken by low-resolution camera on the smartphone under unconstrained background and lighting condition. As the experiments show, the proposed method achieves accurate palm landmark localization within a certain angle of orientation in the wild.

In future work, we plan to collect a significantly larger dataset of palm images and use this dataset to build a better training set to get more accurate results. And apply the palm landmark detection method to a complete palmprint identification and verification system on mobile devices. Furthermore, we will focus on the mathematical theory of SDM to have a deeper analysis of the convergence properties and make it more adaptable to palm landmark detection.

References

[1] M. Ong, T. Connie, and A. Teoh. Michael, "Touch-less palm print biometrics: Novel design and implementation," Jour. Image and Vision Computing, vol. 26, no. 12, pp. 1551-1560, 2008.

[2] M. Choras, R. Kozik, and A. Zelek, "A novel shape-texture approach to palmprint detection and identification," in Eighth International Conference on Intelligent Systems Design and Applications, vol.3, pp. 638-643, 2008.

[3] J. Doublet, O. Lepetit, and M. Revenu, "Contact less hand recognition using shape and texture features," in International Conference on Signal Processing, vol.8, no. 3, pp. 1-4, 2006.

[4] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models: their training and application," Computer Vision Image Understand, vol. 61, no. 1, pp. 38-59, 1995.

[5] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.6, pp. 681-685, 2001.

[6] M. Aykut, M. Ekinci, "AAM-based palm segmentation in unrestricted backgrounds and various postures," Pattern Recognition Letters, vol. 34, no. 9, pp. 955-962, 2013.

[7] X. Xiong, F. De la Torre, "Supervised Descent Method and its Applications to Face Alignment," in IEEE Conference on Computer Vision and Pattern Recognition, pp. 532-539, 2013.

DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-291

[8] A. Kong, D. Zhang, and M. Kamel, "A survey of palmprint recognition, " Pattern Recognition, vol. 42, no.7, pp. 1408-1418, 2009.

[9] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International journal of computer vision, vol. 60, no. 2, pp. 91-110, 2004.

[10] A. Kumar, C. M. Wong, C. Helen Shen, and K. J. Anil, "Personal Verification using Palmprint and Hand Geometry Biometric," in Audio- and Video-Based Biometric Person Authentication, pp. 668-678, 2003.

[11] B. Ozkan, M. Ekinci, and A. Gokdogan, "A new approach stereo based palmprint extraction in unrestricted postures," in IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications Proceedings, pp. 44-49, 2014.