Static Hand Gesture Recognition Using Combinational Features and Compressive Sensing

Huiwei Zhuang, Mingqiang Yang, Zhenxing Cui, Feng Wang

***Abstract*—Vision based hand gesture recognition is becoming a hot research field. This paper proposes a novel static hand gesture recognition method based on combinational features. The method has two new points. First, it combines the gesture's local binary pattern (LBP) feature with curvature scale space (CSS) corner feature together. Second, it employs the compressive sensing (CS) to classify gestures. Through doing experiments on a hand gesture database confirms that recognition rate can reach 96.25%, it performs better than some other methods.**

***Index Terms*—Hand Gesture Recognition, Local Binary Pattern (LBP), Curvature Scale Space (CSS) Corner, Compressive Sensing (CS)**

I. INTRODUCTION

ALONG with the developing of human computer interaction (HCI), more and more researchers have paid attention to the field of hand gesture recognition. Current computer interaction methods with mouse and keyboard are inefficient and inconvenient . As we all know, people from different countries can communicate with each other using hand gestures. If we can also use hand gestures to control computers, it will be more convenient. Fortunately, it is becoming true. Hand gesture recognition can be applied in many fields, such as smart TV, virtual reality, intelligent robot etc [1].

Generally speaking, hand gesture recognition includes data-glove based and vision based. In [2], the author developed a system using a motion tracker and a data glove, the system used strain gauges in the glove to obtain the hand data. But the extra devices are expensive and inconvenient. Comparatively speaking, vision-based gesture recognition does not need any devices on hand, we only need a camera to collect gesture images. When we make some gestures, the machines can make corresponding responds. But this method is not perfect now [3]. This paper only focuses on the static vision based gesture recognition.

Many approaches on gesture recognition have been proposed. The method of recognizing gestures by the signal that carries information on the activity of the fingers was proposed in [4]. But its performance is not good when it is used in the challenging operating conditions. In [5] the author employs the principal components analysis (PCA) to reduce the Gabor filter features, then uses the SVM to recognize. This method can't adapt to the real environment.

In recent years, compressive sensing (CS) is becoming more and more popular, but not many people apply it to recognize gestures. In this paper, we propose a static gesture recognition system based on the CS theory. In view of single feature can't well perfectly express gestures, we combine global features with local features together. Firstly, we do some preprocessing on gesture images, in this step we mainly want to get the ideal hand silhouettes. Secondly, two kinds of gesture features will be extracted including the local binary pattern (LBP) and the curvature scale space (CSS) corner. Lastly, we use the CS to classify gestures. The architecture of the system is shown in Fig 1.

The rest of the paper is organized as following: Section II briefly introduces the preprocessing of original images and describes the LBP feature and the CSS corner feature of the gestures in detail. Section III presents the CS theory and how to combine the two kinds of features to classify gestures. Section IV shows the experimental results. Finally, we draw a conclusion in Section V.

II. PREPROCESSING AND FEATURE EXTRACTION

A. *Preprocessing*

The original images are RGB images, because we want to extract the CSS corners, we must get hand silhouettes.

Firstly, we turn the original images to gray images. Then we use the binarization method based on threshold to get binary images[6]. In binary image the hand region is white, others are black. The hand region at this step we get is defective, because some noise may be produced at the processing procedure. Next, we do morphological processing for several times on binary images. Opening and closing are two kinds of common morphological processing. After the above operating, the hand region is smooth, many small holds and narrow breaks are filled[7]. At last, we use the Laplacian-of-Gaussian (LOG) edge detection method to get hand silhouettes. Fig 2 shows some preprocessing images.

Manuscript received December 21,2015;revised December 30,2015.This work is supported by Shandong Provincial Natural Science Foundation, China (ZR2014FM030, ZR2013FM32 and ZR2014FM010).

Huiwei Zhuang is with the School of Information Science and Engineering, Shandong University, Jinan 250100, China (e-mail: zhuanghw@mail.sdu.edu.cn).

Mingqiang Yang is with the School of Information Science and Engineering, Shandong University, Jinan 250100, China (corresponding author, e-mail: yangmq@sdu.edu.cn, imageinstitute@outlook.com).

Zhenxing Cui is with the School of Information Science and Engineering, Shandong University, Jinan 250100, China (e-mail: 478715631@qq.com).

Feng Wang is with the Shandong Institute for Product Quality Inspection, Jinan 250100, China (e-mail: wfzj0531@sina.com).

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

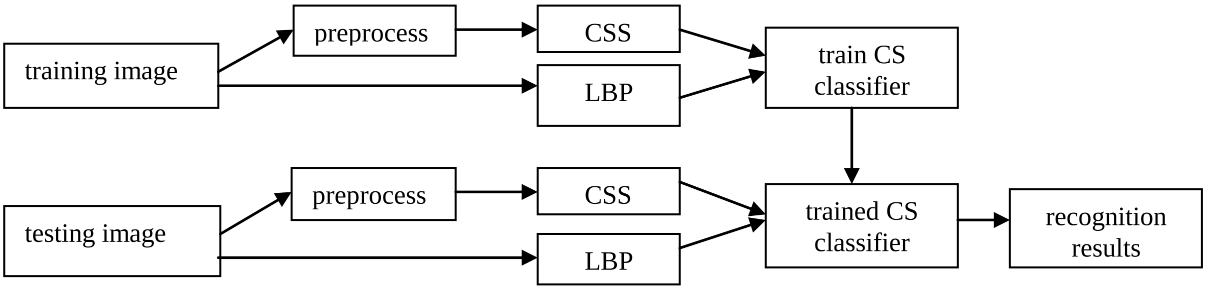
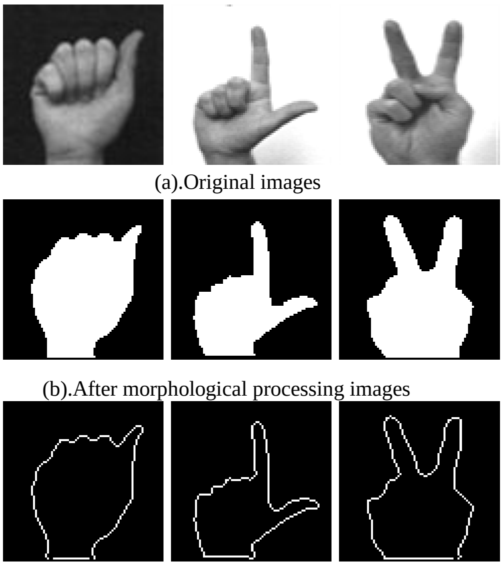


Fig 1. The architecture of the system

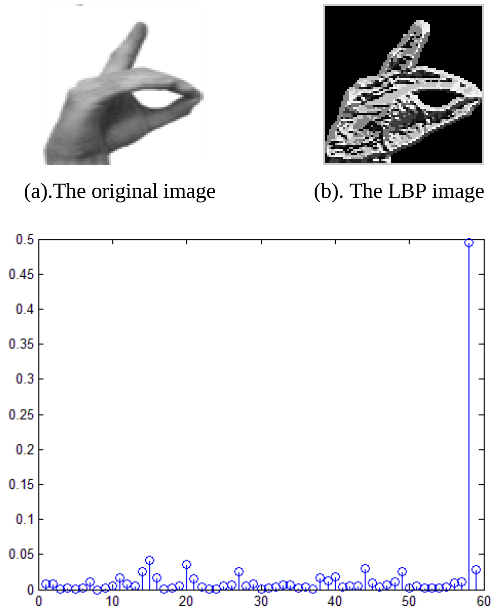


(a).Original images

(b).After morphological processing images

(c). Hand silhouettes

Fig 2. Preprocessing images



(a).The original image (b). The LBP image

(c). The uniform LBP histogram

Fig 3. LBP processing procedure

B. *LBP Feature*

LBP is a kind of local texture feature describing operation of gray images. In the 3×3 neighborhood of every pixel, compare the 8 neighborhoods with the central pixel, if the neighborhood is smaller than the central, this neighborhood is set to be 1, or else it is 0. After the above operation, arrange the 8 binary numbers based on anticlockwise order. Then change it to decimal number. This decimal number is the LBP code of the central pixel. According to the above method, every pixel of the gray image will get a LBP code. After operating on every pixel of a gray image, we will get a LBP code image. In practice, we often use the histogram of the LBP code image to be the feature vector. After the classical LBP algorithm being proposed, some improved methods have been presented, such as the multiresolution LBP, uniform LBP etc [8].

In order to fit different scale texture feature, Timo Ojala *et al* proposed the multiresolution LBP in [9]. Let P be the number and R be the radius of neighborhood, so the LBP operation is expressed as $LBP_{P\text{-}R}$. Some common used are $LBP_{4\text{-}1}$, $LBP_{8\text{-}1}$, $LBP_{8\text{-}2}$, $LBP_{16\text{-}2}$ etc. In this paper we use the $LBP_{8\text{-}1}$.

Along with the P and R being large, the LBP code will increase rapidly. In [9], the author also proposed a kind of uniform LBP to decrease the LBP code. The uniform LBP is a LBP if it only has no more than two times change between 0 and 1. Fig 3 shows the original image, the $LBP_{8\text{-}1}$ image and the uniform LBP histogram.

C. *CSS Feature*

The CSS proposed by F.Mokhtarian *et al* is a kind of feature extraction method based on multi-scale space theory. It extracts the contour shape feature. Firstly, it uses the Gaussian filter $g(\mu,\sigma)$ with different standard deviation σ to filter the contour shape $\Gamma(\mu)=(x(\mu), y(\mu))$. So we will get contour shapes under different scales $\Gamma_\sigma(\mu)$, $\Gamma_\sigma(\mu) = (X(\mu,\sigma), Y(\mu,\sigma))$, where $X(\mu,\sigma)= x(\mu) * g(\mu,\sigma)$, $Y(\mu,\sigma)= y(\mu) * g(\mu,\sigma)$, * is the convolution operation. After we get the contour shapes under different scales, we will compute the curvature at a high scale for every contour shape. The computation of curvature is

$$\kappa(\mu,\sigma)=\frac{X_\mu(\mu,\sigma)Y_{\mu\mu}(\mu,\sigma)-X_{\mu\mu}(\mu,\sigma)Y_\mu(\mu,\sigma)}{(X_\mu(\mu,\sigma)^2+Y_\mu(\mu,\sigma)^2)^{\frac{3}{2}}} \qquad (1)$$

where $X_\mu(\mu,\sigma)=\frac{\partial}{\partial\mu}X(\mu,\sigma)=x(\mu)*g_\mu(\mu,\sigma)$

$$Y_\mu(\mu,\sigma)=\frac{\partial}{\partial\mu}Y(\mu,\sigma)=y(\mu)*g_\mu(\mu,\sigma)$$

$$X_{\mu\mu}(\mu,\sigma)=\frac{\partial^2}{\partial\mu^2}X(\mu,\sigma)=x(\mu)*g_{\mu\mu}(\mu,\sigma)$$

$$Y_{\mu\mu}(\mu,\sigma)=\frac{\partial^2}{\partial\mu^2}Y(\mu,\sigma)=y(\mu)*g_{\mu\mu}(\mu,\sigma)$$

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Then, we compare the local maxima of $\kappa(\mu,\sigma)$ to a single threshold and neighbor minima to get the original corners. At last, we track these corners to the lowest scale and remove those close corners to get the accurate corners[10].

By now, many improved algorithms of CSS have been proposed. In [11], after getting all the original corner candidates by computing the local maxima of $\kappa(\mu,\sigma)$, the author proposed two methods to remove false corners from these candidates. First, he uses an adaptive local curvature threshold instead of a single one to remove rounded corners. The threshold is set based on the neighborhood region's curvature. If the candidate's local maxima is less than its threshold, it will be considered as a false corner and removed. During computing the adaptive threshold, there is a coefficient C, the author has proved that $1<C<2$. In this paper, we take C=1.3. Second, he uses a dynamic region of support (ROS) to check the angles of candidates. In this step, the author defines a maximum obtuse angle that a correct corner can have. In this paper, we take it as 170°. If the angle of a candidate is between 170° and 190°, it is a false corner. Fig 4 shows some corner images generated by the improved CSS. The images are 80×80 size.

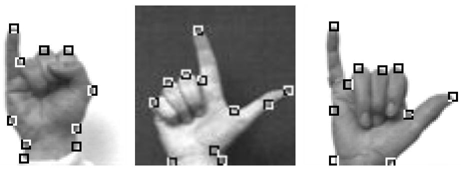


Fig 4. Some corner images of hand

III. CS Theory And Classification

CS is becoming more and more popular in the field of pattern recognition. In this paper, we will introduce the CS theory to the hand gesture recognition.

A. CS Theory

CS theory was established by [12], [13] etc. The theory points that if a signal is sparse, we can reconstruct it from a small number of sample data.

Suppose the original signal is $x \in R^{N\times 1}$. Generally speaking, x is not sparse, but it is sparse in a certain transform domain. $\Psi=[\Psi_1,\Psi_2,\ldots\ldots,\Psi_N] \in R^{N\times N}$ is the sparse basis, x can be represented as $x=\Psi\theta$. $\theta \in R^{N\times 1}$ is the coefficient matrix, it is the sparse representation of x in the transform domain Ψ. If there are only K(K<<N) nonzero coefficients in θ, we call θ is K-parse.

Then we select a measurement matrix $\Phi \in R^{M\times N}$ to measure the original signal x, and get a measurement vector $y \in R^{M\times 1}$, $y=\Phi x$. We can reconstruct x by y. In order to reconstruct the signal x precisely, Φ must satisfy some restricted conditions: (i) $M\geq K\log(N/K)$; (ii) the inner product of Φ and Ψ must satisfy the restricted isometry property (RIP). The measurement matrix formed by independent identically distributed Gaussian variables is used widely[14], [15].

From $x=\Psi\theta$ and $y=\Phi x$, we can get

$$y=\Phi x=\Phi\Psi\theta=A\theta \tag{2}$$

where $A=\Phi\Psi$ is called metric matrix. Considering the noise jamming, (2) can be written as

$$y=A\theta+e \tag{3}$$

where $\|e\|_2 < \varepsilon$, ε is a small constant. In order to reconstruct x, we must solve (3) to get θ, but (3) is underdetermined. This problem can be solved by l1 norm optimization:

$$\theta = \arg\min\|\theta\|_1,\ s.t.\|y - A\theta\|_2 \le \varepsilon \tag{4}$$

After getting θ, from $x=\Psi\theta$, we can get the reconstructed x.

B. Classification

In this paper, we combine LBP with CSS corner features together and employ the CS to classify gestures. The classification steps based on the CS theory are as following:

Step 1: Normalize all images to 80×80, then extract LBP and CSS corner features of all images to get feature vectors.

Step 2: Get the training sparse basis matrix Ψ^{LBP} and Ψ^{CSS} respectively. Supposing, among all training images there are J kinds of gestures and n images of each kind. Then
$\Psi^{LBP}=[\Psi^{LBP}{}_1,\Psi^{LBP}{}_2,\ldots\ldots,\Psi^{LBP}{}_J]$,
$\Psi^{CSS}=[\Psi^{CSS}{}_1,\Psi^{CSS}{}_2,\ldots\ldots,\Psi^{CSS}{}_J]$,
Where $\Psi^{LBP}{}_i=[\Psi^{LBP}{}_{i1},\Psi^{LBP}{}_{i2},\ldots\ldots,\Psi^{LBP}{}_{in}]$, $\Psi^{LBP}{}_{ij}$ is the LBP feature vector of the jth training image of the ith gesture. Similarly, $\Psi^{CSS}{}_i = [\Psi^{CSS}{}_{i1},\Psi^{CSS}{}_{i2},\ldots\ldots, \Psi^{CSS}{}_{in}]$, $\Psi^{CSS}{}_{ij}$ is the CSS corner feature vector of the jth training image of the ith gesture.

Step 3: Take the Gaussian stochastic matrix to be the measurement matrix Φ. On the one hand, use $A^{LBP}=\Phi\Psi^{LBP}$, $A^{CSS}=\Phi\Psi^{CSS}$ to measure sparse basis matrix Ψ^{LBP} and Ψ^{CSS} by Φ respectively to get the metric matrix A^{LBP} and A^{CSS}. On the other hand, supposing x^{LBP} and x^{CSS} are the LBP and CSS feature vectors of a testing image. Use $y^{LBP}=\Phi x^{LBP}$ and $y^{CSS}=\Phi x^{CSS}$ to measure x^{LBP} and x^{CSS} respectively to get the observation set y^{LBP} and y^{CSS}.

Step 4: Make the testing observation set to be the linear combination of the metric matrix by using $y^{LBP}= A^{LBP}\theta^{LBP}+ e$, $y^{CSS}= A^{CSS}\theta^{CSS}+ e$.

Step 5: Use (4) i.e. l1 norm optimization to get the coefficient matrixes θ^{LBP} and θ^{CSS} respectively.

Step 6: Reconstruct the ith testing observation set $y^{LBP}{}_i$ and $y^{CSS}{}_i$ by the relevant optimal coefficient $\theta^{LBP}{}_i$ and $\theta^{CSS}{}_i$ using $y^{LBP}{}_i=\Psi^{LBP}{}_i\theta^{LBP}{}_i$, $y^{CSS}{}_i = \Psi^{CSS}{}_i\theta^{CSS}{}_i$, $i=1,2,\ldots\ldots J$.

Step 7: Combine the two features by two different weights ω_{LBP} and ω_{CSS}. Then use (5) to compute the error between reconstructed testing observation set and the original observation set, the category with the smallest error is the testing image belonging to.

$$i^* = \arg\min_i(\omega_{LBP}\left\|y^{LBP} - y_i^{LBP}\right\|_2^2 + \omega_{CSS}\left\|y^{CSS} - y_i^{CSS}\right\|_2^2) \tag{5}$$

In order to make these steps clearly to see, Fig 5 shows the flow diagram.

IV. Experimental Results

To test the performance of our method, we do some experiments on the Jochen Triesh Database [16]. 720 images taken by 24 different persons in three different backgrounds consist of 10 hand gestures. In this paper, we only use the gestures under two kinds of simple backgrounds. In order to compare with [17], we divide the database into two parts. Select 40 images of every gesture against light and dark background to be the training set,400 in all. The remaining 80

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

images are the testing set. Before doing experiment, we crop all these images to be 80×80 size.

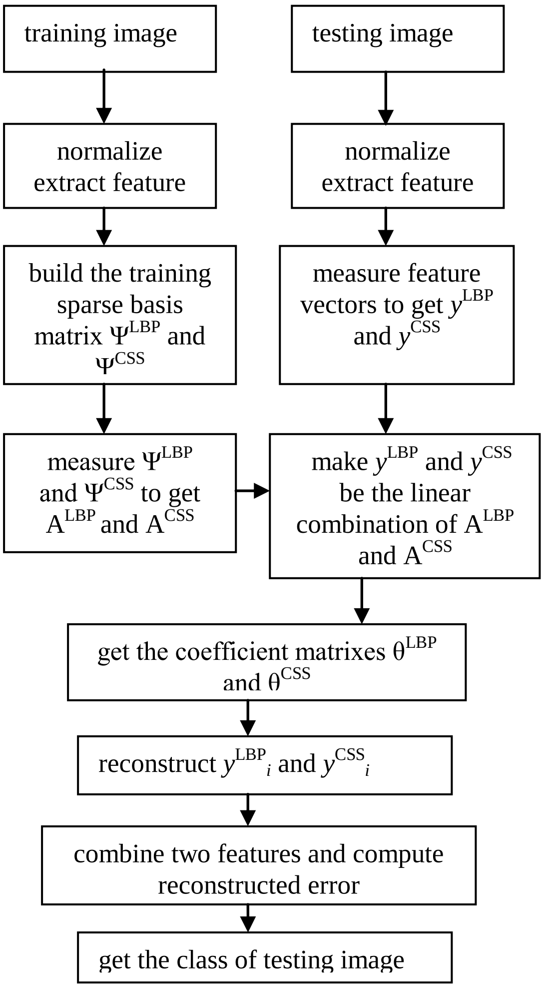


Fig 5. The flow diagram of CS classification

A. Experiment 1

When we use the CS to classify gestures, we combine two kinds of features with different weights ω_{LBP} and ω_{CSS} in (5). Different combination of ω_{LBP} and ω_{CSS} may influent the recognition rate largely. Fig 6 shows the relation between recognition rate and ω_{LBP}. Because $\omega_{LBP} + \omega_{CSS}=1$, here we only consider ω_{LBP}.

From the figure we can see when $\omega_{LBP}=0.63$, $\omega_{CSS}=0.37$ the recognition rate reaches the highest 96.25%. Because $\omega_{LBP}>\omega_{CSS}$, the LBP feature has a more important effect on gesture recognition.

B. Experiment 2

LBP feature is translation invariant, rotation invariant, robust to illumination, and it has the property of simple computation. But it is not scale invariant and it can't express the local features completely. The CSS corner is not only translation invariant, rotation invariant, but also scale invariant. It is a global feature and is robust to noise and occultation. But it is complex of computation and is easy to loss or mistake detecting corners. In order to make up these two kinds methods' shortcomings, we combine these two features together. Fig 7 shows the comparison of the recognition rate of using single feature with combinational features. From the figure we can see that combinational features can improve the recognition performance.

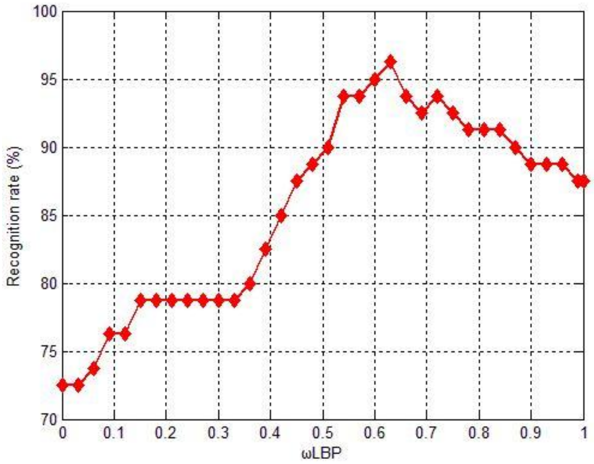


Fig 6. The figure of recognition rate and ω_{LBP}

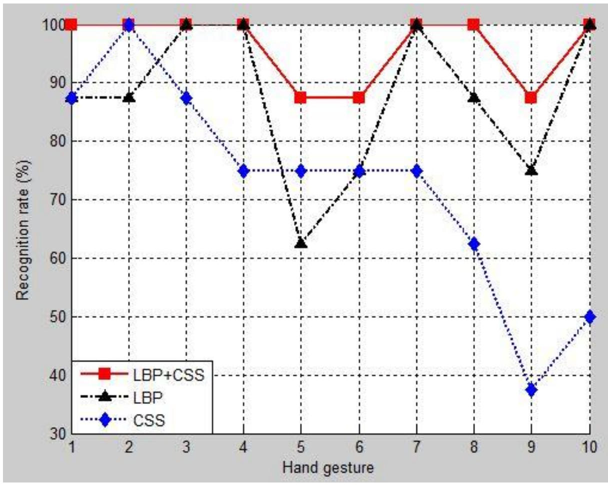


Fig 7. The figure of recognition rate of three methods

C. Experiment 3

This experiment will compare our method with some popular gesture recognition methods recently. In [17], the author uses the Zernike moment and HOG descriptors as the combinational features to recognize. We compare the accuracy between ours with its in Table I. From the table we can see our method has a higher accuracy.

TABLE I
RECOGNITION ACCURACY OF SOME METHODS

Method	Accuracy
[17]	92.1%
Ours	96.25%

V. CONCLUSION

In this paper, we have proposed a new method for vision based hand gesture recognition. In order to make up the

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

shortages of single feature, we combine local feature LBP with global feature CSS corner together. It can overcome some ill effects, such as translation, rotation, scale and so on. Another new point is we employ CS to be the classifier. Some experiments have been done on a hand gesture database. The result which achieves 96.25% accuracy confirms that our method has better recognition performance than traditional methods. In the future, we will pay more attention to increasing the running speed so as to improve the real-time performance of the system.

REFERENCES

[1] P.Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," in World Academy of Science, Engineering and Technology 49,2009,pp.972-977.

[2] Cemil Oz and Ming C. Leu, "Human-computer interaction system with artificial neural network using motion tracker and data glove," in pattern recognition and machine intelligence, 2005, pp.280-286.

[3] Lingchen Chen, Feng Wang, Hui Deng, and Kaifan Ji ,"A survey on hand gesture recognition,"in 2013 International Conference on Computer Sciences and Applications, December 2013, pp.313-316.

[4] A. Malima, E. Özgur, and M. Çetin, "A fast algorithm for vision-based hand gesture recognition for robot control," in Signal Processing and Communications Applications, April 2006, pp.1-4.

[5] D.Y. Huang, W.C.Hu, and S.H. Chang, "Vision-based hand gesture recognition using PCA+Gabor filters and SVM," in 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, September 2009, pp.1-4.

[6] Nobuyuki Otsu, "A threshold selection method from gray-level histogram," in IEEE Transactions on Systerms,Man,and Cybernetics, January 1979, pp.62-66.

[7] D. K. Ghosh and S. Ari, "A static hand gesture recognition algorithm using K-Mean based radial basis function neural network," in 2011 8th International Conference on Information, Communications and Signal Processing (ICICS), December 2011, pp.1-5.

[8] K. Meena and Dr. A. Suruliandi, "Local binary patterns and its variants for face recognition," in IEEE-International Conference on Recent Trends in Information Technology (ICRTIT), June 2011, pp.782-786.

[9] T. Ojala, M. Pietikäinen, Senior Member, IEEE, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," in IEEE Transactions on Pattern Analysis and Machine Intelligence, July 2002, pp.971-987.

[10] F. Mokhtarian and R. Suomela "Robust image corner detection through curvature scale space," in IEEE Transactions on Pattern Analysis and Machine Intelligence, December 1998, pp.1376-1381.

[11] X.C. He and N.H.C. Yung, "Curvature scale space corner detector with adaptive threshold and dynamic region of support," in Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04) , 23-26 August 2004, pp.791-794 Vol.2.

[12] E. J. Candès, J. Romberg, Member, IEEE, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," in IEEE Transactions on Information Theory, February 2006, pp.489-509.

[13] D. L. Donoho, Member, IEEE , "Compressed sensing," in IEEE Transactions on Information Theory, April 2006, pp.1289-1306.

[14] V. M. Patel and R. Chellappa "Sparse representations, compressive sensing and dictionaries for pattern recognition," in 2011 First Asian Conference on Pattern Recognition (ACPR), November 2011, pp.325-329.

[15] P. Nagesh and B. X. Li, "A compressive sensing approach for expression-invariant face recognition," in IEEE Conference on Computer Vision and Pattern Recognition,June 2009, pp.1518-1525.

[16] J. Triesch, C. Malsburg, S.Marcel, Hand posture and gesture datasets: Jochen Triesch static hand posture database, Available at http://www.idiap.ch/resources/gestures.

[17] H. L. Zhang, H. Y. Li, and M. Zhou, "Hand posture recognition based on multi-feature and compressive sensing," in Journal of Hunan University (Natural Sciences), March 2013, pp.87-92.

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)