Gesture recognition based on sparse representation

Wei Miao, Gongfa Li, Ying Sun*,
Guozhang Jiang and Jianyi Kong

College of Machinery and Automation,
Wuhan University of Science and Technology,
Wuhan 430081, China
Email: 962698619@qq.com
Email: ligongfa@wust.edu.cn
Email: 493530316@qq.com
Email: whjgz@wust.edu.cn
Email: 15697188659@wo.com.cn
*Corresponding author

Honghai Liu

State Key Laboratory of Mechanical System and Vibration,
School of Mechanical Engineering,
Shanghai Jiao Tong University,
Shanghai 200240, China
and
Intelligent Systems and Biomedical Robotics Group,
School of Computing,
University of Portsmouth,
Portsmouth, PO1 3HE, UK
Email: honghai.liu@sjtu.edu.cn
Email: honghai.liu@port.ac.uk

Abstract: Aiming at the problem that the robustness of gesture recognition is difficult to guarantee, this paper presents a method based on multi-features and sparse representation. Hu invariant moments and HOG features of training samples are extracted in training phase. The K-SVD algorithm is used to train the initial value of dictionary formed by two features so as to obtain two sub-dictionaries. In recognition phase, sparse coefficients of corresponding training dictionary are derived by solving minimum l_1-norm. Finally, the overall reconstruction error is calculated to judge the categories of test samples. In experimental simulation, five kinds of grasp gesture are collected to create gesture sample library. After selecting optimal HOG parameters and the weight of two features, the recognition effect of the method is analysed. Compared with the commonly used classification, the results show that the method has better recognition rate and robustness.

Keywords: human–computer interaction; gesture recognition; Hu invariant moments; HOG feature; sparse representation.

Reference to this paper should be made as follows: Miao, W., Li, G., Sun, Y., Jiang, G., Kong, J. and Liu, H. (2016) 'Gesture recognition based on sparse representation', *Int. J. Wireless and Mobile Computing*, Vol. 11, No. 4, pp.348–356.

Biographical notes: Wei Miao received BS degree in Mechanical Engineering and Automation from Zhengzhou Huaxin College, Zhengzhou, China. He is currently occupied in his MS degree in Mechanical Design and Theory at Wuhan University of Science and Technology. His current research interests include mechanical CAD/CAE, signal analysis and processing.

Gongfa Li received the PhD degree in Wuhan University of Science and Technology, Wuhan, China. He is currently a Professor in Wuhan University of Science and Technology. His major research interests are computer-aided engineering, mechanical CAD/CAE, modelling and optimal control of complex industrial process.

Ying Sun is currently an Associate Professor in Wuhan University of Science and Technology. Her major research focuses on teaching research in mechanical engineering.

Guozhang Jiang received the PhD degree in Wuhan University of Science and Technology, China. He is currently a Professor in Wuhan University of Science and Technology. His research interests are computer-aided engineering, mechanical CAD/CAE and industrial engineering and management system.

Jianyi Kong received the PhD degree in Helmut Schmidt Universitat, Germany. He is currently a Professor in Wuhan University of Science and Technology. His research interests are intelligent machine and controlled mechanism, mechanical and dynamic design and fault diagnosis of electrical system, mechanical CAD/CAE, intelligent design and control.

Honghai Liu received the PhD degree in intelligent robotics from Kings College, University of London, London, UK. He is currently a Professor of School of Mechanical Engineering in Shanghai Jiao Tong University, Shanghai, China, and a Professor of Intelligent Systems in Portsmouth University, Portsmouth, UK. His research interests are approximate computation, pattern recognition, multi-sensor-based information fusion and analytics, human–machine systems, advanced control, intelligent robotics and their practical applications.

1 Introduction

In recent years, with the development of mechanical automation and the diversification of operation task, a type of multi-finger dexterous hand become the development trend of the robot end actuators, and the research of anthropomorphic prosthetic hand also provides convenience for the disabled people's normal life (Liu et al., 2016; Ciancio et al., 2016). In order to have similar operation capacity of human hand, it is necessary to adopt the human–computer interaction way to transfer human skills, and gesture recognition is one of the hot research problems in the field of human–computer interaction. Although gesture recognition based on vision technology has made great development, it still has drawbacks, including recognition rate, robustness, real-time performance, stability and practicability and so on. Currently, the gesture itself has the characteristics of uncertainty and diversity, such as inaccurate changes of the hand shape, occlusion, movement speed, etc., and there are often difficulties such as complex background environment, illumination uncertain conditions in the real environment, the change of image space information and computer processing power, so that the study of gesture recognition based on vision has certain challenges (Yeo et al., 2013; Belsare and Sujatha, 2015).

Feature extraction and classification recognition are two important steps in gesture recognition. In terms of feature extraction, statistical moment features, such as Hu moment and Zernike moment, have the characteristics of scale invariance, translation invariance and rotation invariance. Luo et al. (2012) added three expressions as the features of gesture recognition on the basis of Hu moment. Fourier descriptor has a good ability to describe the outline (Luo et al., 2012). Ren and Zhang (2009) adopted 12 Fourier descriptors as the feature vector. Histogram of Oriented Gradient (HOG) is formed as feature through calculating the local area's histogram, which was used for pedestrian detection initially and for gesture recognition in recent years (Ren and Gu, 2011). A variety of features can be extracted from the gesture images; however, how to combine the various features to improve the recognition rate is a problem that needs to be studied. Classification algorithms include template matching, Support Vector Machine (SVM), neural network, etc. Ren and Zhang (2009) developed an improved MEB-SVM algorithm on the basis of SVM, which reduced the computational complexity. Hasan and Kareem (2014) predicted ten static gestures with NN classifier. Yang and Sun (2014) put forward a kind of gesture recognition algorithm based on the BP neural network of quantum particle swarm optimisation to improve the learning efficiency of BP neural network. Despite there are many gesture recognition methods, traditional classifiers require a complex training process and are sensitive to noise.

Sparse representation has become a hot research spot in the field of pattern recognition and computer vision in recent years, which provides a new perspective to solve the classification problem (Wang et al., 2014; Zhao and Yang, 2015; Yu and Fang, 2016). Sparse representation, which is different from the traditional Nyquist sampling theorem, maps the high-dimensional signal to low-dimensional space by training an over-complete dictionary that is not related to the transformation matrix and reconstructs the original from a small amount of signal after obtaining sparse vector (Cui and Prasad, 2015). When the training samples are enough, the test samples can be represented linearly by the same type of training sample and other kind of samples' contribution to refactor the test sample is zero so as to better describe the similarity degree information between samples. Sparse representation classification has been successfully applied to face recognition for the first time by Wright et al. (2009), who verified the SRC method has good robustness in the aspect of face recognition. Sparse representation classification needs to be sparse-represent the features received from the targets, namely a complete dictionary based on the features of targets is trained, which linear represents the test samples and reconstructs samples, and the objects are recognised according to minimum residual error between the test samples and reconstruction samples (Ortiz and Becker, 2014).

In this paper, the sparse representation theory is applied to the gesture recognition by constructing sparse representation classifier. Firstly, two features of samples are extracted as the initial value of the dictionary, and then some features are selected to generate a new dictionary using K-Singular Value Decomposition (K-SVD). Finally, the objects' category is determined through test objects' sparse degree in dictionary and reconstruction residual. Experimental results show that recognition effect of the method is better than other commonly used methods in different rotation angle, size and illumination condition.

2 Sparse representation theory

Sparse representation refers to that the images (or features) are completely or approximately linear combination of a group of atoms that very few images (or features) and all the atomic images (or features) form an over-complete dictionary (Boyali and Kavakli, 2012; Srinivas et al., 2013). Given n_k training samples from k-th class as columns of a matrix $A_k = \left[v_{k1}, v_{k2}, \cdots, v_{kn_k}\right] \in \Re^{m \times n_k}$. Training set has c classes and $A = \left[A_1, A_2, \cdots, A_c\right]$ are all training sets. New test sample $y \in \Re^m$ from k-th class can be represented as a linear combination of one type of elements:

$$\hat{x} = \arg\min_x \|x\|_0 \ \ s.t.\ y = Ax$$

If $n_k \leq m$, the equations $y = Ax$ is overdetermined and x has unique solution. However, $y = Ax$ is underdetermined in gesture recognition, so it has infinitely many solutions. Generally, this difficulty is resolved by minimum l_0-norm.

$$\hat{x} = \arg\min_x \|x\|_0 \ \ s.t.\ y = Ax$$

But the problem of finding the solution of l_0-norm is NP-hard, so approximate solutions are adopted generally. Algorithms of research include: greedy algorithm, including matching tracking algorithm, orthogonal matching pursuit algorithm, etc., to realise signal approximation through choosing appropriate atoms and a series of progressively increasing method (Boyali and Hashimoto, 2016). Convex optimisation algorithm, including gradient projection method, tracking methods the minimum point of regression method, etc., put the l_0-norm relax to l_1-norm by solving linear programming (Khan and Raja, 2016). Convex optimisation algorithm is more accurate than greedy algorithm, but requires higher computational complexity.

According to the principle of compression sensing, when coefficients are sparse enough, the difficulty can be converted into l_1-norm:

$$\hat{x} = \arg\min_x \|x\|_1 \ \ s.t.\ y = Ax$$

Because data generally contains noise, the constraints can have a certain error range:

$$\hat{x} = \arg\min_x \|x\|_1 \ \ s.t.\ y = Ax$$

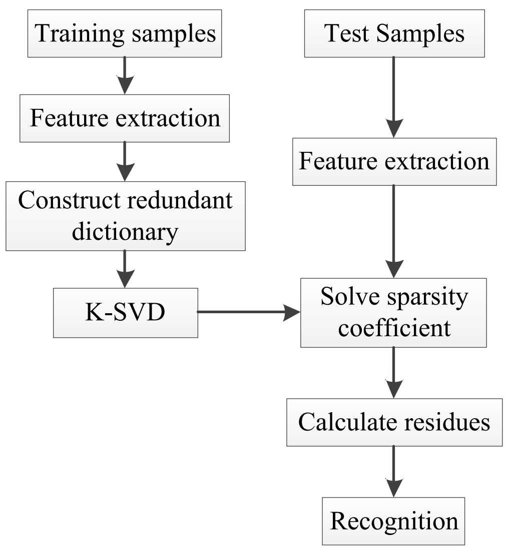
The dictionary is the key to sparse representation of the original signal. Dictionary is generally constructed by two methods: one way is generated directly from the training sample. Because it has similar characteristics with test samples, test signal can be more easily represented by such dictionary (Guo et al., 2016). Another method is to train training samples for a better dictionary. The atoms of dictionary are few but can represent the whole sample, which can reduce the amount of calculation in classification. The commonly used algorithms are MOD algorithm and K-SVD algorithm (Sun and Wang, 2014).

Today, the sparse representation theory has been widely used, such as image restoration, image de-noising, image classification, image recognition, etc. One of the most noteworthy is sparse representation classification (SRC), and test samples are assigned to the class of minimum refactoring error constructed by the training samples.

3 Identification framework based on sparse representation

Gesture recognition is used to obtain image signal of gesture by image sensor, preprocess, extract features and classify or identify (Chen et al., 2014). Tradition object recognition used the optimal hyper-plane, template matching and weak classifiers, which have certain classification ability, but are unable to get a better effect from the angle of sparse. In this paper, the proposed method is conducted in the segmentation image, extracts Hu invariant moments and HOG features as the initial value of redundant dictionary, adopts the K-SVD algorithm for training, and finally classifies by sparse classification. The recognition framework is shown in Figure 1.

Figure 1 Identification framework



3.1 HOG feature

HOG is a feature descriptor used for object detection in computer vision and image processing. Dalal and Triggs (2005) put forward HOG for the first time, which is used for the pedestrian detection in static image or video. It calculates statistics image gradient direction histogram of local area to form characteristics, and the shape of the local object can be described well by density distribution of gradient or edge direction (Vanbang et al., 2015; Xu et al., 2016). HOG features can keep certain robustness under different illumination and scale. The extraction process of HOG feature is as follows:

1 Normalise the global image. In order to reduce the influence of illumination factors, the whole image needs to be normalised. Each colour channel calculates the square root, respectively, and gamma is 0.5.

$$I(x,y) = I(x,y)^{gamma}$$

2 Calculate the gradient of image. The horizontal gradient is obtained by convolution operation between the gradient operator [–1, 0, 1] and original image. The vertical gradient is obtained by convolution operation between the gradient operator $[-1, 0, 1]^T$ and original image.

$$Gx(x,y) = I(x+1,y) - I(x-1,y)$$
$$Gy(x,y) = I(x,y+1) - I(x,y-1)$$
$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2}$$
$$\alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right)$$

where $I(x,y)$ is pixel value, $G_x(x,y)$ is horizontal gradient, $G_y(x,y)$ is vertical gradient, $G(x,y)$ is gradient magnitude, and $\alpha(x,y)$ is gradient orientation.

3 Construct HOG for each cell. The images are divided into multiple cells, which are not overlap. The gradient of each pixel in each cell is calculated. The gradient orientation can be 0°–360° or 0°–180°, and 0°–180° is better. The gradient orientation is divided into nine bins, and weighted projection is for each bin using the gradient of each cell, so that HOG of each cell is obtained.

4 Normalise the gradient histogram in block combined with cells. Combining multiple adjacent cells into a block, and the feature has better invariance for light, shadow and edge contrast after normalising each block's gradient. The normalised function is l_1-norm and l_2-norm generally, and Dalal's article proves that l_2-norm has good effect.

5 Generate feature description vector. Features, collected in all the overlapping blocks, are combined into the final feature vectors for classification.

3.2 Hu invariant moments

Moment in statistics is used to reflect the distribution of random variables. If we take image grey value as a 2D or 3D density distribution function, the moment can be used for the field of image analysis and the extraction of image features (Lin et al., 2015; Fernando and Wijjayanayake, 2015). M.K. Hu proved that Hu invariant moments have the characteristics of scale invariance, translation invariance and rotation invariance. Hu invariant moments can make up for the inadequacy of HOG features to some extent, its calculation process is:

1 For discrete digital image, image function is $f(x,y)$, and the $p + q$ order geometric moment and central moments of image are:

$$m_{pq} = \sum_{y=1}^{N}\sum_{x=1}^{M} x^p y^q I(x,y) \quad p,q = 0,1,2,\cdots$$

$$\mu_{pq} = \sum_{y=1}^{N}\sum_{x=1}^{M} (x-\bar{x})^p (y-\bar{y})^q I(x,y) \quad p,q = 0,1,2,\cdots$$

where N and M are the height and width of the image, respectively, $\bar{x}$ and $\bar{y}$ represent centroid of the image, $\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}}$.

2 μ_{pq} has translation invariance, but still sensitive to rotate. Normalised centre distance is defined as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}{}^{\rho}} \quad \rho = (p+q)/2+1$$

3 The second order and third order normalised central moments are constructed seven moment invariants M_1–M_7:

$$M_1 = \eta_{20} + \eta_{02}$$
$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}{}^2$$
$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} + \eta_{03})^2$$
$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$
$$M_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right)$$
$$M_6 = (\eta_{20} - \eta_{02})\left((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$
$$M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right) - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})\left(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right)$$

3.3 Recognition based on sparse representation

The sparse representation abilities of dictionaries trained by different classes of training samples are very different, and the test samples are more likely to be represented by the dictionary trained by same classes of training samples (Zhou et al., 2013; Bomma et al., 2013). According to the characteristics of the sparse representation, the classifier based on sparse representation is designed.

Given c classes gesture data and n_k (k = 1, 2, …, c) training sample of k-th class. HOG feature vector is G_k and Hu invariant moments vector is H_k, and the corresponding eigenvectors of a test samples are y_1 and y_2. We initialise two dictionaries $D_1 = [G_1, G_2, \cdots, G_k]$ and $D_2 = [H_1, H_2, \cdots, H_k]$, and get redundant dictionary trained by K-SVD algorithm, which can use less but optimal data to represent the whole training set and reduce the calculation of classification.

The sparse representation coefficients $\hat{x}$ of training samples can be expressed as the linear combination between the sparse representation coefficients of each feature and feature weight. x_i is i-th feature's sparse representation coefficients, w_i is i-th feature's weight, $\sum_{i=1}^{2} w_i = 1$.

$$\hat{x} = w_1 x_1 + w_2 x_2 = \sum_{i=1}^{2} w_i x_i$$

The test data can be represented by two dictionaries D_i, and describe two features' contribution for classification by adjusting w_i. The problem can be described as l_1-norm problem:

$$\hat{x} = \arg\min_{x_i} \left(w_1 \|x_1\| + w_2 \|x_2\| \right) = \arg\min_{x_i} \left(\sum_{i=1}^{2} w_i \|x_i\|_1 \right)$$
$$s.t. \quad \sum_{i=1}^{2} y_i = \sum_{i=1}^{2} D_i x_i$$

Because the data often have noise, the optimisation problem is changed into:

$$\hat{x} = \arg\min_{x_i} \left(w_1 \|x_1\| + w_2 \|x_2\| \right) = \arg\min_{x_i} \left(\sum_{i=1}^{2} w_i \|x_i\|_1 \right)$$
$$s.t. \quad \sum_{i=1}^{2} \left(D_i x_i - y_i \right) \le \varepsilon$$

It is equivalent to $\hat{x} = \arg\min_{x_i} \left(\sum_{i=1}^{2} \|D_i x_i - y_i\|_2^2 + \lambda \|w_i x_i\|_1 \right)$.

Finally, the overall minimum reconstruction error of a single class is calculated to determine the category of the test samples:

$$\hat{k} = \arg\min_{k} \sum_{i=1}^{2} \left\| y_i - D_i \delta_k(\hat{x}) \right\|_2$$

where $\delta_k(\hat{x})$ is same dimension with $\hat{x}$, and only keep corresponding element of $\hat{x}$ with k-th class, namely $\delta_k(\hat{x}) = \left[0, 0, \cdots, x_{k1}, \cdots, x_{k2}, \cdots, x_{kn_k}, \cdots, 0, 0\right]$.

To sum up, the classification algorithm based on multi-features and sparse representation is summarised as follows:

1 Input: a matrix of training samples $D_1 = [G_1, G_2, \cdots, G_k]$ and $D_2 = [H_1, H_2, \cdots, H_k]$ for k classes, a test sample $Y = [y_1, y_2]$, and error tolerance $\varepsilon > 0$.

2 Normalise the columns of training samples and train dictionary by K-SVD algorithm.

3 Solve sparse representation coefficients of different sub-dictionary, namely the l_1-norm problem:

$$\hat{x} = \arg\min_{x_i} \left(w_1 \|x_1\| + w_2 \|x_2\| \right) = \arg\min_{x_i} \left(\sum_{i=1}^{2} w_i \|x_i\|_1 \right)$$
$$s.t. \quad \sum_{i=1}^{2} \left(D_i x_i - y_i \right) \le \varepsilon$$

4 Compute the residuals $\gamma_k = \sum_{i=1}^{2} \left\| y_i - D_i \delta_k(\hat{x}) \right\|_2$

5 Output: $\hat{y} = \arg\min_{k} \gamma_k$

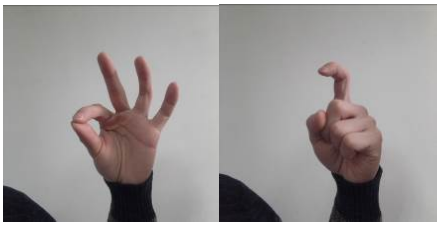
4 Experimental simulation

4.1 Data acquisition

In order to verify the recognition ability of this method, we need to collect data, set up the sample library and analyse the influence of each factor on gesture recognition. As shown in Figure 2, five kinds of grasp gesture samples are collected from different people, corresponding to grasping movements of multi-fingered hand, including open, grasp by five fingers, grasp by three fingers, pinch and hook. Figure 3 shows the gesture samples, which are collected in different rotation angle, different size and different illumination condition. Each class of gestures has 100 training samples and 50 test samples. All images only contain gestures and the background is white to ensure extracted features are all from the gesture area after segmentation and accuracy of training dictionary. Figure 3 (d) shows the images after adding noise, which took as the samples to verify the robustness of the algorithm. The ordinary camera is used for gesture acquisition and its image resolution is 640 × 480. The hardware configuration of computer in experiment is Intel Core i5 CPU and 4G memory. Software environment is Windows 7 operating system. Simulation platform is MatlabR2010b. In order to recognise more easily and intuitively, the algorithm is encapsulated into a function and GUI of Matlab is used to compile the gesture recognition system, as shown in Figure 4. In the system of *Hand Recognition via SRC*, the training set images are selected with the button *Train*; the single test gesture image is selected for identification with the button *SingleTest*, and its category is outputted; the multi-test gesture images are selected with the button *MultiTest*, and the recognition rate is outputted.

Figure 2 Five gesture samples

A B C



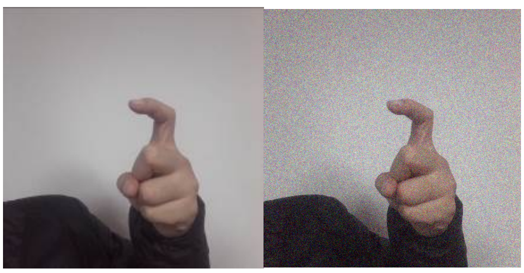
D E

Figure 3 Collected sample images

(a) Different size

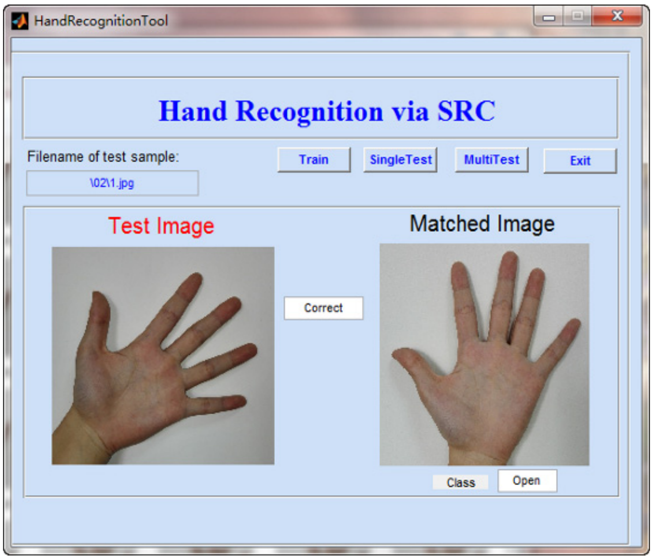
(b) Different rotation angle

(c) Different illumination conditions



(d) Noised images

Figure 4 GUI recognition result



4.2 Parameter selection

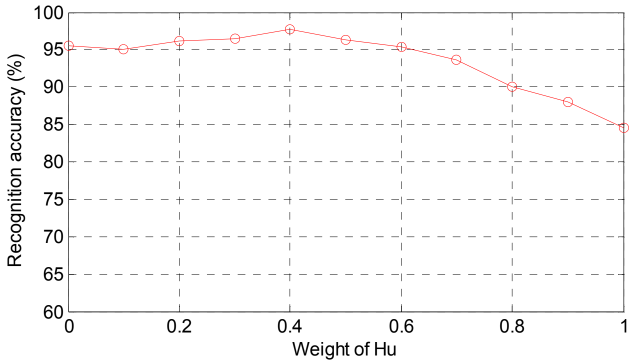
After pretreatment for all the samples, we need to choose some parameters when extract the HOG feature: size of window $ws \times ws$, size of each block $bs \times bs$, size of the overlap between blocks ol and size of each cell in the block $cs \times cs$ (Xu et al., 2014). Range of gradient direction is 0°–180°, which is divided into nine bins averagely. The more cells in block and overlap between the blocks, the more dimension for HOG. The recognition rate is calculated in this paper by testing some parameter combinations so as to find the optimal parameter combination. Seen from Table 1, the recognition rate of $ws42\_bs18\_ol10\_cs6$ is higher.

Table 1 Recognition rate of different parameter combinations of HOG

ws	*bs*	*ol*	*cs*	*Dimension*	*Recognition rate (%)*
36	12	3	6	900	96.6
36	18	9	6	729	96.9
40	16	8	8	576	96.9
40	16	4	8	324	95.8
40	16	8	4	2304	97.5
40	16	4	4	1296	97.7
42	18	6	6	729	96.8
42	18	10	6	1296	98.6

After selecting the optimal combination parameters of HOG, we need to test the recognition effect of the weights of Hu moments and HOG features in different values of Zhang et al. (2013). We set $\omega_1 + \omega_2 = 1$. When $\omega_1 = 1$, the recognition rate is expressed only using the Hu moment. When $\omega_1 = 0$, the recognition rate is expressed only using the HOG feature. Seen from Figure 5, when the weight of Hu moments is $\omega_1 = 0.4$ and the weight of HOG features $\omega_2 = 0.6$, the recognition rate is higher. From the influence of the value of ω_1 and ω_2 for recognition performance can be seen, the impact of different features for the results of hand gesture recognition is different, and the influence level of HOG features is greater than the Hu moments features.

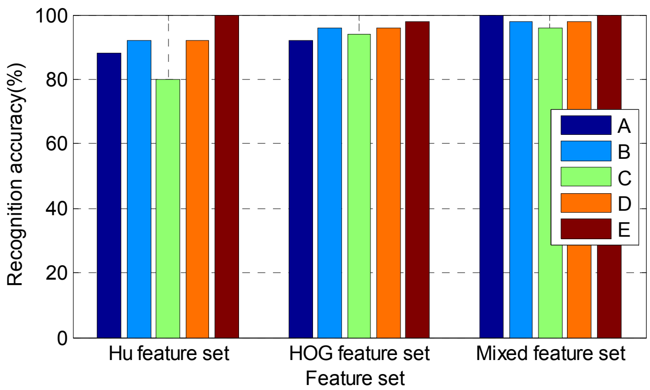
Figure 5 Recognition rate of different weight (see online version for colours)



4.3 *Recognition effect*

After collecting hand gesture samples, we select the optimal HOG feature parameters and two feature weights, extract HOG features and Hu invariant moments of the samples and recognise by the sparse representation classifier. Figure 6 displays the gestures recognition rate on different feature sets, the recognition rate of two kinds of feature fusion is high. Although there is no obvious difference with HOG feature, recognition of rotation gesture is improved after combining Hu moments. Table 2 is the confusion matrix on different feature sets. Because of the impact of image binarysation, there are many wrong gestures of Hu moment set in Table 2. The improvement of binary effect can improve the recognition rate. In Table 2, the recognition effect of the HOG feature set is better than *t* Hu moment set, but it is not correct to identify the rotating samples. The recognition rate in Table 2 is better, because two features complement each other after using HOG feature to extract the local features and Hu moments feature to extract the global feature.

Figure 6 Recognition rate on different feature sets



4.4 *Robustness*

Different noises are superimposed on the image to test, such as Gaussian noise, Poisson noise and salt and pepper noise, the algorithm can still correctly identify. The algorithm also shows certain robustness in different scale, rotation and light. On the one hand, the training samples contain gestures under all conditions. On the other hand, the adopted features describe the gestures better from the global and local perspective. HOG feature eliminates the influence of illumination and small angle rotation, and Hu moment is a rotation invariant. The combination of the two features further improves the overall recognition performance. Commonly used classification methods, including SVM, Back Propagation Neural Network (BPNN) and *K*-Nearest Neighbour (*K*NN), were tested in the condition of this paper. Figure 7 is the recognition rate under the different conditions, which shows that the algorithm shows strong advantages in terms of robustness compared with the commonly used method of gesture recognition.

Table 2 Confusion matrix on different feature sets

(a) Confusion matrix on Hu moment feature set

Gesture	*A*	*B*	*C*	*D*	*E*
A	44	6	1	0	0
B	3	46	1	0	0
C	0	5	40	5	0
D	0	2	2	46	0
E	0	0	0	0	50

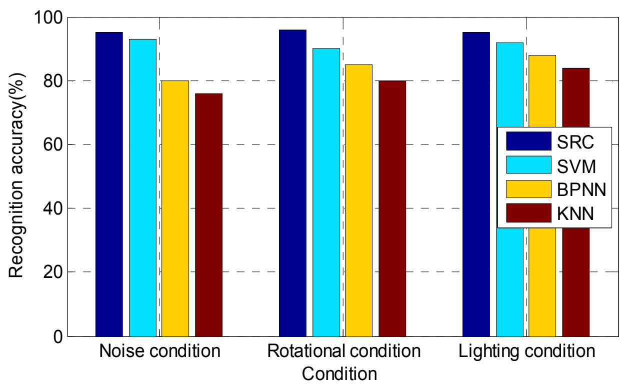
(b) Confusion matrix on HOG feature set

Gesture	*A*	*B*	*C*	*D*	*E*
A	46	4	0	0	0
B	0	48	2	0	0
C	1	2	47	0	0
D	0	0	2	48	0
E	1	0	0	0	49

(c) Confusion matrix on fusion feature set

Gesture	*A*	*B*	*C*	*D*	*E*
A	50	0	0	0	0
B	0	49	1	0	0
C	0	2	48	0	0
D	0	0	1	49	0
E	0	0	0	0	50

Figure 7 Recognition rates of different methods under different conditions



5 Conclusion and future work

In order to improve the accuracy and robustness of gesture recognition, the paper applies sparse representation to gesture recognition. Firstly, the hand gestures under different conditions are collected. The optimal parameters of HOG feature and the weights of two features are selected. Then, Hu invariant moment and HOG feature are extracted. The K-SVD dictionary training method is used to choose some atoms representing all features to reduce the computation cost. Finally, sparse representation classifier is constructed for identification. Compared with other methods, the recognition rate of this method is higher. The next step of research will add some shaded image and improve the speed of sparse representation algorithm.

Acknowledgement

This work was supported by grants of National Natural Science Foundation of China (Grant Nos. 51575407, 51575338, 1575412, 61273106).

References

Belsare, S. and Sujatha, K. (2015) 'Robust part-based hand gesture recognition using kinect sensor', *International Journal of Engineering Sciences & Research Technology*, Vol. 4, No. 8, pp.75–80.

Bomma, S., Favaro, P. and Robertson, N.M. (2013) 'Sparse representation based action and gesture recognition', *IEEE International Conference on Image Processing*, IEEE, Melbourne, Australia, pp.141–145.

Boyali, A. and Hashimoto, N. (2016) 'Spectral collaborative representation based classification for hand gestures recognition on electromyography signals', *Biomedical Signal Processing and Control*, Vol. 24, pp.11–18.

Boyali, A. and Kavakli, M. (2012) 'A robust gesture recognition algorithm based on sparse representation, random projections and compressed sensing', *IEEE Conference on Industrial Electronics and Applications*, IEEE, Singapore, pp.243–249.

Chen, S.B., Xu, L.X. and Luo, B. (2014) 'Multiple kernel sparse representation-based classification', *Acta Electronica Sinica*, Vol. 42, No. 9, pp.1807–1811.

Ciancio, A.L., Cordella, F., Barone, R., Romeo, R.A., Dellacasa Bellingegni, A., Sacchetti, R., Davalli, A., Di Pino, G., Ranieri, F., Di Lazzaro, V., Guglielmelli, E. and Zollo, L. (2016) 'Control of prosthetic hands via the peripheral nervous system', *Frontiers in Neuroscience*, Vol. 10, pp.1–17.

Cui, M. and Prasad, S. (2015) 'Class-dependent sparse representation classifier for robust hyperspectral image classification', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 5, pp.2683–2695.

Dalal, N. and Triggs, B. (2005) 'Histograms of oriented gradients for human detection', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, San Diego, CA, pp.886–893.

Fernando, M. and Wijjayanayake, J. (2015) 'Novel approach to use HU moments with image processing techniques for real time sign language communication', *International Journal of Image Processing* Vol. 9, No. 6, pp.335–345.

Guo, Y., Zhao, G. and Pietikäinen, M. (2016) 'Dynamic facial expression recognition with atlas construction and sparse representation', *IEEE Transactions on Image Processing*, Vol. 25, No. 5, pp.1977–1992.

Hasan, H. and Kareem, S.A. (2014) 'Static hand gesture recognition using neural networks', *Artificial Intelligence Review*, Vol. 41, No. 2, pp.147–181.

Khan, R. and Raja, R. (2016) 'Introducing L1-sparse representation classification for facial expression recognition', *Imperial Journal of Interdisciplinary Research*, Vol. 2, No. 4, pp.115–122.

Lin, K., Zhang, Y.D., Dang, Y.W. and Huang, J.H. (2015) 'SAR image matching algorithm base on improved Hu invariant moments', *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 8, No. 1, pp.1–8.

Liu, H., Yang, D., Fan, S. and Cai, H. (2016) 'On the development of intrinsically-actuated, multisensory dexterous robotic hands', *ROBOMECH Journal*, Vol. 3, No. 1, pp.1–9.

Luo, Y., Xie, Y. and Zhang, Y. (2012) 'Design and implementation of a gesture-driven system for intelligent wheelchairs based on the kinect sensor', *Robot*, Vol. 34, No. 1, pp.110–114.

Ortiz, E.G. and Becker, B.C. (2014) 'Face recognition for web-scale datasets', *Computer Vision and Image Understanding*, Vol. 118, No. 1, pp.153–170.

Ren, Y. and Gu, C.C. (2011) 'Hand gesture recognition based on HOG characters and SVM', *Bulletin of Science and Technology*, Vol. 27, No. 2, pp.211–213.

Ren, Y. and Zhang, F.M. (2009) 'Hand gesture recognition based on MEB-SVM', *International Conferences on Embedded Software and Systems*, 25–27 May, Zhejiang, pp.344–349.

Srinivas, U., Chen, Y., Monga, V., Nasrabadi, N.M. and Tran, T.D. (2013) 'Exploiting sparsity in hyperspectral image classification via graphical models', *IEEE Geoscience and Remote Sensing Letters*, Vol. 10, No. 3, pp.505–509.

Sun, R. and Wang, J.J. (2014) 'A vehicle recognition method based on kernel K-SVD and sparse representation', *Pattern Recognition and Artificial Intelligence*, Vol. 27, No. 5, pp.435–442.

Vanbang, L.E., Zhu, Y., Zhao, J.K. and Chen, N. (2015) 'Real-time gesture recognition method based on depth image HOG features', *Journal of East China University of Science and Technology (Natural Science Edition)*, Vol. 41, No. 5, pp.698–702.

Wang, R., Du, L.F., Sun, D. and Wan, W.G. (2014) 'Traffic object recognition in complex scenes based on SIFT and kernel sparse representation', *Acta Electronica Sinica*, Vol. 42, No. 11, pp.2129–2134.

Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S. and Ma, Y. (2009) 'Robust face recognition via sparse representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp.210–227.

Xu, D., Wu, X., Chen, Y.L. and Xu, Y. (2014) 'Online dynamic gesture recognition for human robot interaction', *Journal of Intelligent & Robotic Systems*, Vol. 77, No. 3, pp.583–596.

Xu, Y., Xu, X.L., Li, C.N., Jiang, M. and Zhang, J.G. (2016) 'Pedestrian detection combining with SVM classifier and HOG feature extraction', *Computer Engineering*, Vol. 42, No. 1, pp.56–60.

Yang, Z.Q. and Sun, G. (2014) 'Gesture recognition based on quantum-behaved particle swarm optimization of back propagation neural network', *Journal of Computer Applications*, Vol. 34, No. 1, pp.137–140.

Yeo, H.S., Lee, B.G. and Lim, H. (2013) 'Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware', *Multimedia Tools and Applications*, Vol. 74, No. 8, pp.2687–2715.

Yu, C.J. and Fang, Y.C. (2016) 'Applying sparse representation-based feature selection to face recognition for gender and race', *Computer Applications and Software*, No. 1, pp.138–141.

Zhang, H.L., Li, H.Y. and Zhou, M. (2013) 'Hand posture recognition based on multi-feature and compressive sensing', *Journal of Hunan University (Natural Sciences)*, Vol. 40, No. 3, pp.87–92.

Zhao, Y.Q. and Yang, J. (2015) 'Hyperspectral image denoising via sparse representation and low-rank constraint', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 1, pp.296–308.

Zhou, Y., Liu, K., Carrillo, R.E., Barner, K.E. and Kiamilev, F. (2013) 'Kernel-based sparse representation for gesture recognition', *Pattern Recognition*, Vol. 46, No. 12, pp.3208–3222.