Multimed Tools Appl (2016) 75:6021–6034
DOI 10.1007/s11042-015-2947-0

Hand gesture recognition using saliency and histogram intersection kernel based sparse representation

Wenji Yang[1,2] · **Lingfu Kong**[2] · **Mingyan Wang**[3]

Received: 13 November 2014 / Revised: 19 July 2015 / Accepted: 14 September 2015 /
Published online: 26 September 2015
© Springer Science+Business Media New York 2015

Abstract Nowadays, sparse representation classification (SRC) has been widely applied in various computer vision areas such as face recognition. However, few researchers have applied SRC in static hand gesture recognition. In our paper, we propose to employ saliency based feature and sparse representation for hand gesture recognition and make in-depth researches in sparsity term parameter and sparse coefficient computation. In addition, literatures show that SRC can not deal with non-linear features well and may produce bad recognition results, so we propose to employ histogram intersection kernel function to map the original features into kernel feature space and use sparse representation classification in the kernel feature space. Furthermore, we compare SR with Support Vector Machine (SVM), Artificial Neural Networks (ANN), Bayesian Network (BN) and Decision Tree (DT). At last, experimental results show that the recognition rate obtained using l1ls_featuresign algorithm has a higher recognition rate than that of l1_ls algorithm and Sparse Representation outperforms all other classifiers compared. In addition, the performance comparison on different kernel functions and different features is also conducted. The average recognition rate of saliency based feature on histogram intersection kernel is 98.91 %, indicating the effectiveness of the proposed saliency based feature and the histogram intersection kernel.

Keywords Histogram intersection kernel · Saliency histogram feature · Kernel sparse representation · Hand gesture recognition

✉ Wenji Yang
ywenji614@163.com

✉ Mingyan Wang
mingyanw@ncu.edu.cn

Lingfu Kong
lfkong@ysu.edu.cn

[1] School of Software, Jiangxi Agricultural University, Nanchang, China

[2] School of Information Science and Engineering, Yanshan University, Qinhuangdao, China

[3] School of Information Engineering, Nanchang University, Nanchang, China

1 Introduction

Hand gesture recognition has wide applications such as human-computer interaction, virtual reality, sign language recognition, distance education, rehabilitation training. So far extensive researches have been conducted and a variety of classification approaches have emerged for hand gesture recognition, such as Support Vector Machine (SVM) [3, 19], Artificial Neural Network (ANN) [16], Bayesian Network (BN) [17], Decision Tree (DT) [6], Nearest Neighbor (NN), Template Matching (TM) [22], Discriminative Ferns Ensemble (DFE) [11], Classification via Regression (CR). For example, Wang et al. proposed to extract LBP feature of 2D patch in the disparity cost map and employed SVM for recognition [19]. Kaoning et al. proposed multi-scale topological features for hand gesture presentation, comparison of which against Modified Census Transform (MCT) [7] and Histogram of Oriented Gradient (HOG) [15] features on three different classifiers such as Decision Tree, Bayesian Network and Classification via Regression [6] was conducted. Keskin proposed a novel Random Decision Forest based hand shape classifier and then used it in a novel multi-layered RDF framework for articulated hand pose estimation [8].

In recent years, sparse representation is widely used in many computer vision areas including image compression and denoising [1], face recognition [20, 24], video-based action classification [13] and so on. For instance, in [20], impressive results were obtained in face recognition by using sparse representation classification and showed that occlusion and corruption could be tackled in sparse representation classification framework. In [16], Meng et al. proposed a robust sparse coding (RSC) by modeling sparse coding as a sparsity-constrained robust regression problem for face recognition and extensive experiments demonstrated that the RSC scheme was much more effective than state-of-the-arts. However, only a few efforts have been made to apply this technique in hand gesture recognition. Stergios et al. proposed a sparse representation-based dynamic hand gesture recognition approach, in which hand coordinates at each frame were utilized as features, however, conventional OMP and BP method was employed to obtain sparse coefficients [14]. In addition, Zhou et al. and Xu et al. also proposed to use sparse representation for dynamic gesture recognition [21, 26]. However, those earlier methods all solved for dynamic gesture recognition and used sparse representation for recognition in straightforward ways without in-depth researches in sparse coding methods of sparse representation. Considering the advantages of sparse representation method and its wide applications especially the dynamic hand gesture recognition, we explore its potential in static hand gesture recognition. In this paper, we transform the optimization problem of coefficient learning in sparse representation into L1-regularized least squares problem and then employ the feature-sign search algorithm proposed in efficient sparse coding to solve L1-regularized least squares problem.

Most hand gesture recognition methods utilize features such as local binary pattern, gradient, motion, surf, coordinates and geometric features for recognition. However, saliency has the ability to remove complex background and uniformly highlight objects, suggesting an effective method to characterize hand gestures. Hence we propose to apply saliency based feature in recognition. In addition, literature [5, 25] shows that the use of the original feature space in sparse representation may not produce good results under certain circumstances. Therefore, inspired by the nonlinear generalization ability of kernel function, we propose histogram intersection kernel function based sparse representation in our hand gesture recognition and different kernel functions are compared in experiments.

Fig. 1 Five different kinds of static gestures

2 Saliency based feature construction

Saliency, which describes the degree of stimulation of each pixel to human eyes, is computed through imitating visual attention mechanism of human vision system and fusing different low level features and high level priors. Saliency maps obtained at detection stage can be used as an effective feature to represent hand gesture, hence we propose to use it in our hand gesture recognition study.

In our paper, we employ two approaches to construct saliency feature. One method is to resize the $M \times N$ saliency map into $m \times n$ and transform it to a d-dimensional vector ($d=m*n<M*N$), while the other method is to construct a block radial histogram based saliency feature, namely saliency histogram feature.

We borrow the block radial histogram idea from literature [18] and adapt it to construct block radial histogram based saliency feature for hand gesture recognition. The details are as follows:

Step 1: Compute the saliency using the proposed multi-scale global regional contrast method [23].
Step 2: Divide the normalization window into 2×2 sub-windows.
Step 3: Divide each sub-window into 18 sector zones, each accounting for 20°.
Step 4: Sum all the saliency values in each sector zone and 18-dimensional feature vector is formed. At last, feature vectors from 4 sub-windows are connected into a 4×18=72 dimension feature vector.

Table 1 Hand gesture recognition rate based on l1_ls and l1ls_featuresign algorithm

Extracted feature		Saliency	Surf	gra	lbp
Gesture 1	l1_ls	97.00 %	94.50 %	84.00 %	77.00 %
	l1ls_fs	99.00 %	97.50 %	92.00 %	85.50 %
Gesture 2	l1_ls	95.50 %	97.00 %	91.00 %	86.50 %
	l1ls_fs	96.50 %	99.50 %	97.50 %	88.00 %
Gesture 3	l1_ls	99.50 %	98.00 %	99.50 %	99.00 %
	l1ls_fs	99.00 %	99.50 %	99.50 %	99.00 %
Gesture 4	l1_ls	95.50 %	65.50 %	88.50 %	77.00 %
	l1ls_fs	94.50 %	85.50 %	90.00 %	78.50 %
Gesture 5	l1_ls	98.50 %	96.50 %	99.00 %	99.00 %
	l1ls_fs	98.50 %	99.00 %	97.50 %	99.00 %
Average	l1_ls	97.20 %	90.30 %	92.40 %	87.70 %
	l1ls_fs	97.50 %	96.20 %	95.30 %	90.10 %

Table 2 Effect of different values on recognition rate

	10^{-4}	10^{-3}	10^{-2}	0.1	0.2	0.3	0.4	0
Gesture 1	70 %	80.5 %	83 %	92 %	92 %	90 %	86.5 %	33.5 %
Gesture 2	75 %	83.5 %	92 %	96.5 %	98 %	99.5 %	99 %	46 %
Gesture 3	98 %	99.5 %	99 %	100 %	98 %	95.5 %	96.5 %	98.5 %
Gesture 4	75 %	85 %	93.5 %	91.5 %	88.5 %	83 %	85 %	28.5 %
Gesture 5	99.5 %	97.5 %	99 %	100 %	99 %	100 %	100 %	99 %

3 Sparse representation model

The basic idea of sparse representation is to represent an image of unknown hand gesture label using different types of hand gesture images with known hand gesture labels. By minimize the reconstruction error, the unknown hand gesture could be labeled accordingly. The details are as follows.

3.1 The theory of sparse representation [20]

Supposing there are n images in the training set, which correspond to K types of hand gestures. The feature vector of each image in the training set is $v \in R^m$, the number of images belonging to the i-th class is n_i, where $i=1,2,\ldots,K$. Construct a matrix $A_i = \left[v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}\right] \in R^{m \times n_i}$ using n_i images corresponding to the i-th class. Each column of A_i corresponds to each hand gesture image in the i-th class. Supposing the training samples of the i-th class are sufficient $A_i = \left[v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}\right] \in R^{m \times n_i}$,any sample $y \in R^m$ from the same class will approximately lie in the linear span of the training samples associated with the i-th class, namely

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \ldots + \alpha_{i,n_i} v_{i,n_i} \tag{1}$$

where $\alpha_{i,j} \in \mathbb{R}$, $j=1,2,\ldots,n_i$

Since the membership i of the test sample is initially unknown, we define a new matrix A for the training samples associated with all the classes as the concatenation of n training samples of all K classes:

$$A = [A_1, A_2, \ldots, A_K] = \left[v_{1,1}, v_{1,2}, \ldots v_{1,n_1}, \ldots, v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}, \ldots, v_{K,1}, v_{K,2}, \ldots, v_{K,n_K}\right] \tag{2}$$

Table 3 Recognition rates based on sparse representation and other classifiers

Extracted feature	Gesture 1	Gesture 2	Gesture 3	Gesture 4	Gesture 5	Average
SR	99 %	96.5 %	99 %	94.5 %	98.5 %	97.5 %
SVM	92 %	87.5 %	94.5 %	90 %	90.5 %	90.9 %
ANN	74.5 %	76.5 %	86 %	79.5 %	96.5 %	82.6 %
BN	84.5 %	85 %	88 %	89.5 %	85 %	86.4 %
DT	79.5 %	84 %	93.5 %	81.5 %	94 %	86.5 %

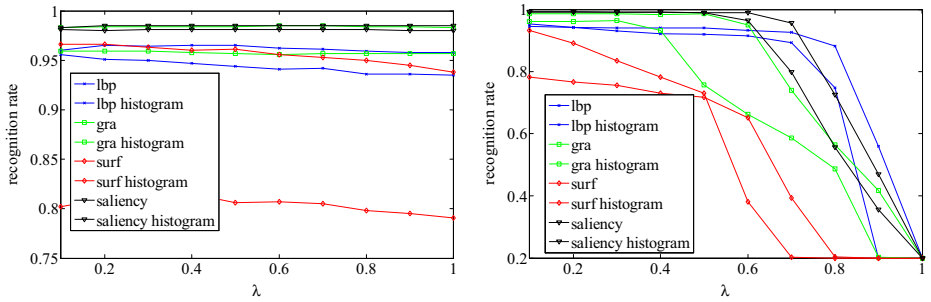

Fig. 2 Recognition rates comparison among different features based on PK (*left*) and IDK (*right*)

where $n_1+n_2+\ldots+n_K=n$. Therefore, the linear representation of y can be rewritten in terms of all the training samples as

$$y = Ax_0 \in \mathbb{R}^m \tag{3}$$

where $x_0 = \left[0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots 0\right]^T \in \mathbb{R}^n$, which is a coefficient vector, only the entries corresponding to the i-th class is nonzero and other entries are all zero.

Therefore, the entries of vector x_0 encode the class which the test sample y attributes to. To obtain it, it needs to solve the linear system of equations.

In the process of linear representation, there usually exists representation error unless the database contains the test image. To solve this problem, we introduce the error limitation, for instance, $\|e\|_2 < \varepsilon$.

$$y = Ax + e \tag{4}$$

The number of images in the database are generally much larger than the dimension of feature vector ($n \gg m$), so the above equations are an underdetermined system of equations. The underdetermined system of equations is an ill-posed problem, which can not directly obtain x from y. The direct solution is to achieve x through solving l_0 -norm minimization problem, l_0 -norm representing the number of nonzero entries in vector x.

$$\min \|x\|_0 \quad s.t. \quad y = Ax \tag{5}$$

l_0 -norm minimization problem belongs to the non-deterministic polynomial time complexity problem. According to the proof in the paper [4]: if $y \in R^m$ is represented by the basis in A

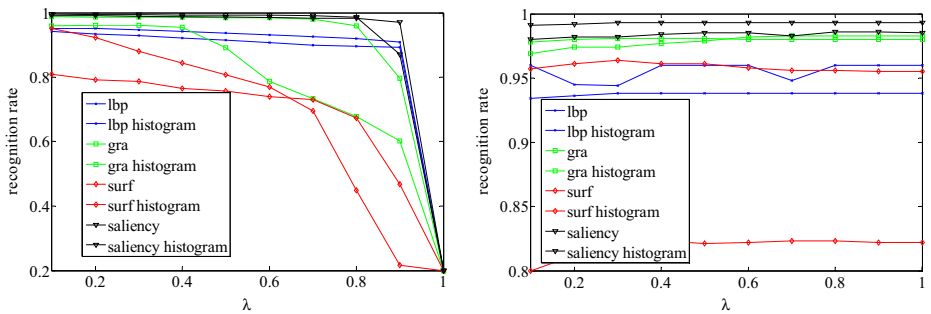

Fig. 3 Recognition rates comparison among different features based on ISDK (*left*) and eHIK (*right*)

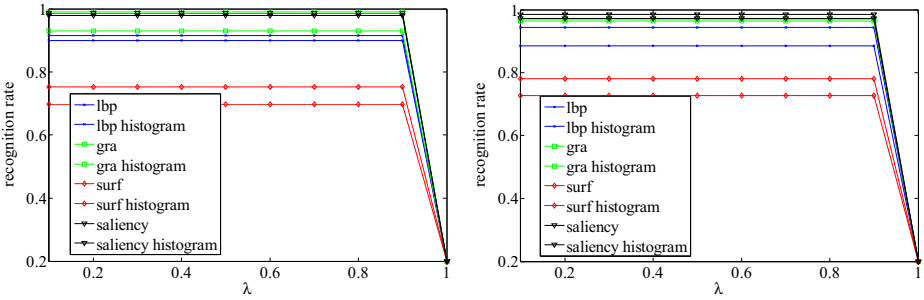

Fig. 4 Recognition rates comparison among different features based on GK (*left*) and EK (*right*)

and $\|y\|_0 < \left(\sqrt{2}-0.5\right)/\rho$ (ρ represents the strength of correlation among basis in A), the solution of l_1 -norm minimization problem can be equivalent to the solution of l_0 -norm minimization problem. Therefore, it can be converted to solve the following l_1 -norm minimization problem:

$$\min\|x\|_1 \quad s.t. \quad y = Ax \tag{6}$$

l_1 -norm minimization problem is a convex optimization problem and can be solved by mathematics [2]. A lot of researchers in recent years solve the optimization problem through obtaining the approximate solution, which is restricted to a certain error conditions

$$\min\|x\|_1 \quad s.t. \quad \|y-Ax\|_2^2 \le \varepsilon \tag{7}$$

3.2 Sparse representation-based hand gesture recognition

In this paper, we employ the feature-sign search algorithm for solving the above l_1 -norm minimization problem [12] and obtain the sparse representation vector of the test sample.

Now the sparse representation vector of the test sample is available, and then we conduct sparse representation-based hand gesture recognition using the obtained sparse representation vector. The detail algorithm is as follows:

(1). Input: Matrix $A=[A_1,A_2,\ldots,A_K]\in\mathbb{R}^{m\times n}$ obtained using all the training samples, test sample $y\in\mathbb{R}^m$, (Allowable errors $\varepsilon>0$).
(2). Normalize each column of A using l_2 -norm.

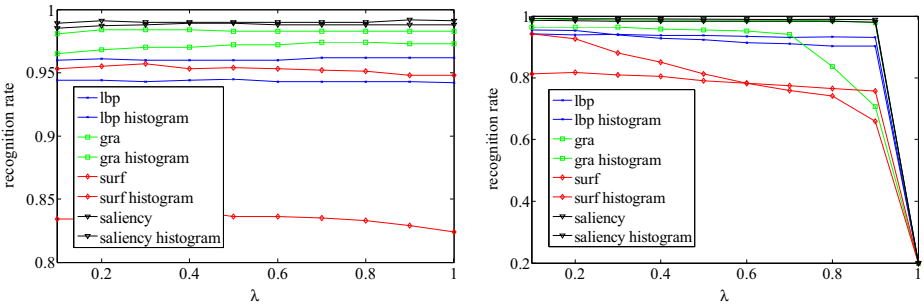

Fig. 5 Recognition rates comparison among different features based on HIK (*left*) and LK (*right*)

Table 4 Hand gesture recognition rates of different kernel functions and different hand gestures based on saliency feature

	Gesture 1	Gesture 2	Gesture 3	Gesture 4	Gesture 5	Average
PK	97.50 %	99.00 %	100.00 %	95.85 %	98.00 %	98.07 %
IDK	99.90 %	66.60 %	76.75 %	63.80 %	84.60 %	78.33 %
ISDK	99.85 %	86.20 %	88.90 %	86.30 %	89.35 %	90.12 %
eHIK	99.40 %	97.35 %	98.10 %	97.10 %	99.95 %	98.38 %
GK	98.65 %	89.10 %	89.10 %	87.30 %	89.55 %	90.74 %
EK	99.55 %	87.75 %	89.55 %	86.85 %	89.55 %	90.65 %
HIK	99.50 %	98.15 %	99.00 %	97.25 %	100.00 %	98.78 %
LK	99.85 %	88.60 %	88.75 %	85.65 %	90.00 %	90.57 %

(3). To solve l_1 -norm minimization problem:

$$\hat{x}_1 = \arg\min_x \|x\|_1 \quad s.t. \quad \|y - Ax\|_2^2 \le \varepsilon$$

(4). Compute residual $r_i(y) = \|y - A\delta_i(\hat{x}_1)\|_2$, where $i=1,2,\ldots,K$, $\delta_i(\hat{x}_1)$ represent a vector which is formed by setting the entries in $\hat{x}_1$ which correspond to samples, not belonging to the i-th class, to zero.

(5). Output:if $\arg\min_i r_i(y)$ exceeds a certain threshold e, it is considered that hand gesture y is an unknown gesture. Then add hand gesture y with a certain amount number of samples to the training set A and retrain using the new training set. Otherwise the class hand gesture y belongs to is $\arg\min_i r_i(y)$.

4 Histogram intersection kernel function based sparse representation

From the view of l_1 -minimization algorithm, the atoms associated with different classes in dictionary must be distinguishable or separable. However, in practice, the atoms obtained in the original feature space may overlap with each other, which consequently produces poor classification results [26]. Because sparse representation employs unit l_2 -norm to normalize training samples and uses the normalized training samples as dictionary atoms, which causes l_1

Table 5 Hand gesture recognition rates of different kernel functions and different hand gestures based on saliency histogram feature

	Gesture 1	Gesture 2	Gesture 3	Gesture 4	Gesture 5	Average
PK	98.50 %	99.00 %	100.00 %	95.40 %	99.50 %	98.48 %
IDK	99.35 %	73.80 %	82.75 %	71.80 %	87.15 %	82.97 %
ISDK	98.95 %	89.25 %	89.20 %	86.60 %	89.55 %	90.71 %
eHIK	99.50 %	99.95 %	99.50 %	97.90 %	99.50 %	99.27 %
GK	99.55 %	88.20 %	87.75 %	85.50 %	89.55 %	90.11 %
EK	99.10 %	87.30 %	89.10 %	82.80 %	89.55 %	89.57 %
HIK	99.50 %	98.65 %	100.00 %	97.00 %	100.00 %	99.03 %
LK	99.45 %	89.10 %	90.00 %	87.15 %	90.00 %	91.14 %

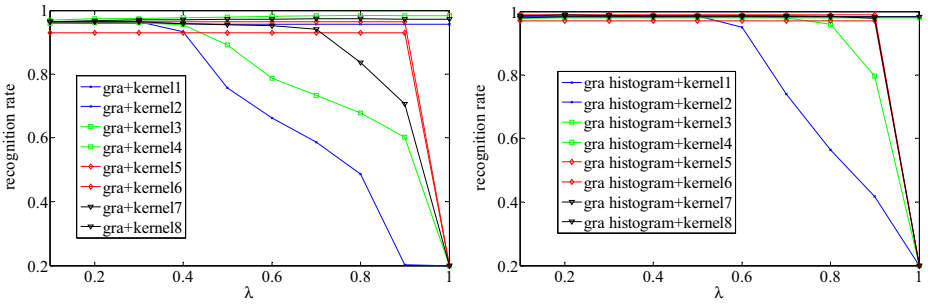

Fig. 6 Recognition rates comparison among different kernel functions on gradient (*left*) and gradient histogram feature (*right*)

-minimization algorithm confusion in selecting the true atoms. In addition, sparse representation cannot solve non-linear features well. Kernels are widely used in machine learning and can transform nonlinear problems into linear problems. Therefore, inspired by the nonlinear generalization ability of kernel function, we employ kernel sparse representation in hand gesture recognition.

Kernel sparse representation [5] is a process to map the original feature space into a high dimensional feature space and conducts sparse representation in high dimensional feature space. Supposing there exists kernel function $k(.,.)$, which deduced by $\varphi: \mathbb{R}^m \to \mathbb{R}^d$, where $m \ll d$, $k(u_i, u_j)=\varphi(u_i)\cdot\varphi(u_j)$ represents the nonlinear similarity between vector $u_i \in R^m$ and vector $u_j \in R^m$, the function maps the input feature $y \in R^m$ and basis $A \in R^{m\times n}$ into high dimensional feature space:

$$y \in R^m \xrightarrow{\varphi} \varphi(y) \in R^d \tag{8}$$

$$\begin{aligned} A &= \left[v_{1,1}, v_{1,2}, \ldots v_{1,n_1}, \ldots, v_{K,1}, v_{K,2}, \ldots, v_{K,n_K}\right] \xrightarrow{\varphi} \Phi \\ &= \left[\varphi\left(v_{1,1}\right), \varphi\left(v_{1,2}\right), \ldots \varphi\left(v_{1,n_1}\right), \ldots, \varphi\left(v_{K,1}\right), \varphi\left(v_{K,2}\right), \ldots, \varphi\left(v_{K,n_K}\right)\right] \end{aligned} \tag{9}$$

where $n_1+n_2+\ldots+n_K=n$, $\Phi \in R^{d\times n}$

For the convenience of description, here we rewrite the above formulation as:

$$A = [u_1, u_2, u_3, \ldots u_n] \xrightarrow{\varphi} \Phi = [\varphi(u_1), \varphi(u_2), \varphi(u_3), \ldots \varphi(u_n)] \tag{10}$$

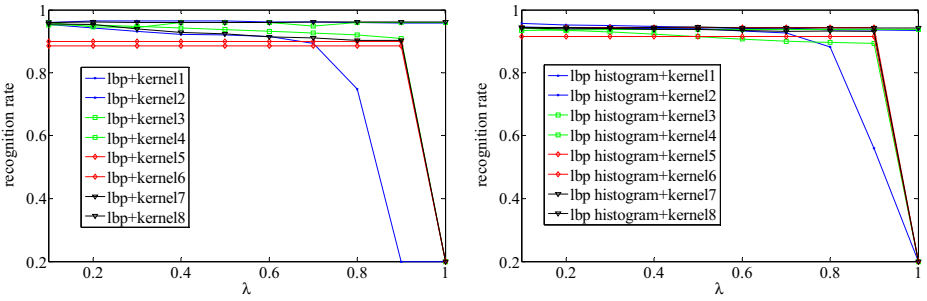

Fig. 7 Recognition rates comparison among different kernel functions on lbp (*left*) and lbp histogram feature (*right*)

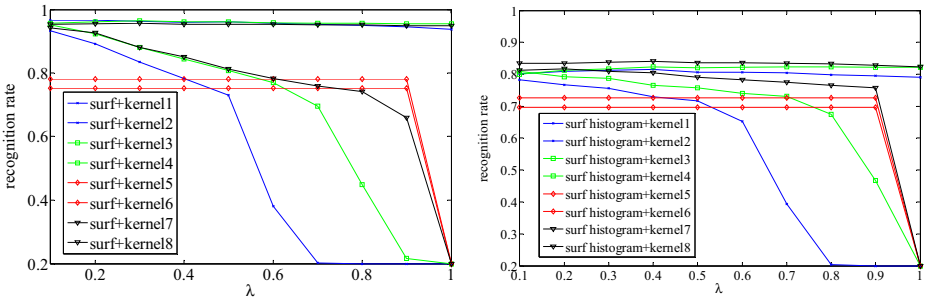

Fig. 8 Recognition rates comparison among different kernel functions on surf (*left*) and surf histogram feature (*right*)

Therefore, we can obtain kernel sparse representation model as follows:

$$\min \|x\|_1 \quad s.t. \quad \|\varphi(y) - \Phi x\|_2^2 \le \varepsilon \tag{11}$$

Through introducing Lagrange multiplier λ, formulation (11) can be transformed as follows:

$$\min_x \lambda \|x\|_1 + \|\varphi(y) - \Phi x\|_2^2 \tag{12}$$

where parameter λ is a weight to balance the sparsity and the reconstruction error.

Assume K_{AA} is a $n \times n$ matrix, where (i, j) entry $K_{AA}(i,j)=k(u_i,u_j)$, K_{Ay} is a n dimension vector, where the i-th entry $K_{Ay}(i)=k(u_i,y)$, so above formulation can be rewritten as:

$$\min_{A,x} \lambda \|x\|_1 + k(y,y) + x^T K_{AA} x - 2x^T K_{Ay} \tag{13}$$

Set $L(x)=k(y,y)+x^T K_{AA} x-2x^T K_{Ay}$. The method for solving the above formulation is the same as solving the common sparse representation model, except that the definitions of K_{AA} and K_{Ay} are different, which can be defined using any kernel functions. When linear kernel function is used in kernel sparse representation, namely $K_{AA}=A^T A$ and $K_{Ay}=A^T y$, then kernel sparse representation is transformed into sparse representation.

In our paper, we employ histogram intersection kernel function to compute K_{AA} and K_{Ay}. Here we fix codebook A, so we can optimize sparse coding using the proposed kernel based version of feature-sign search algorithm [5]. The details are as follows:

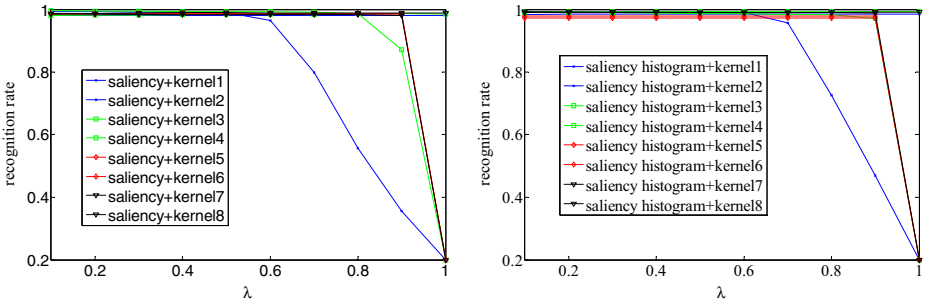

Fig. 9 Recognition rates comparison among different kernel functions on saliency (*left*) and saliency histogram feature (*right*)

Input: Codebook A, input feature y and the weight of sparsity term λ.
Output: Kernel sparse code x.

Step 1 Initialize $x = \overrightarrow{0}$, $\theta = \overrightarrow{0}$, *activeset*={}, where $\theta_i \in \{-1,0,-1\}$ is the sign of the i-th entry x^i of vector x.

Step 2: Compute the followings using histogram intersection kernel function.

$$L(x) = k(y,y) + x^T K_{AA} x - 2x^T K_{Ay},$$
$$L_x = \frac{\partial L(x)}{\partial x} = 2\left(K_{AA} x - K_{Ay}\right)$$
$$L_{xx} = \frac{\partial^2 L(x)}{\partial x^2} = 2K_{AA}$$

Step 3: From zero coefficients of x, select $k=\arg\max_i |L_x^i|$.
If $L_x^k > \lambda$, then $\theta_k = -1$, *activeset*=$\{k\} \cup$*activeset*.
If $L_x^k < -\lambda$, then $\theta_k = 1$, *activeset*=$\{k\} \cup$*activeset*.

Step 4: Feature-sign step

Step 4.1: Let $\hat{A}$ and $\hat{L}_{xx}$ be sub-matrix of A and L_{xx} which contain only the columns corresponding to *activeset*, let $\hat{x}$, $\hat{K}_{Ay}$ and $\hat{\theta}$ be sub-vectors of x, K_{Ay} and θ corresponding to *activeset*.

Step 4.2: Obtain the solution of the resulting unconstrained QP: $\min_{\hat{x}} L(\hat{x}) + \lambda \hat{\theta}^T \hat{x}$, the solution is:

$$\hat{x}_{new} = \left(\hat{L}_{xx}\right)^{-1}\left(2\hat{K}_{Ay} - \lambda\hat{\theta}\right)$$

Step 4.3: Perform a discrete line search on the closed segment from $\hat{x}$ to $\hat{x}_{new}$.
Check the objective value at $\hat{x}_{new}$ and all points where any coefficient changes sign.
Update $\hat{x}$ to the point with the lowest objective value.

Step 4.4: Remove zero coefficients from *activeset* and update $\theta = sign(x)$.

Step 5: Check the optimality conditions:

(1) Optimality condition for nonzero coefficients: $L_x^j + \lambda sign(x^j) = 0, \forall x^j \neq 0$.
If condition (1) is not satisfied, go to Step4; else check condition (2).

(2) Optimality condition for zero coefficients: $|L_x^j| \leqslant \lambda$, $\forall x^j = 0$.
If condition (2) is not satisfied, go to Step3, else return x as the solution.
After sparse coefficients are obtained, we can use sparse representation-based hand gesture recognition method to perform kernel sparse representation-based hand gesture recognition.

5 Experiments and analysis

In our paper, we use benchmark Cambridge hand gesture dataset in our experiments. Five types of static hand shape and total 13,461 hand gesture images are selected based on criteria of illumination changes, uneven illuminations and shadows in images. 400

static images are randomly selected for training and testing, in which one half is used for training and the rest is for testing. Five types of static gestures are demonstrated in Fig. 1.

5.1 Comparison between different algorithms for sparse coefficient computation

In our experiments, saliency, surf, gradient and lbp features are used for sparse representation-based hand gesture recognition. In addition, we employ l1_ls algorithm implemented by Kwangmoo Koh et al. [9, 10] and l1ls_featuresign algorithm implemented by Bruno Olshausen et al. [12] to solve l1-normalized least square problems for sparse coefficient computation. In l1_ls algorithm, λ=0.1, τ=0.8, $\tau \in (0,1]$, c=8. In l1ls_featuresign algorithm, β=0.4.

The recognition rates using above two algorithms are demonstrated in Table 1, which indicates a higher average recognition rates of l1ls_featuresign algorithm than l1_ls algorithm and a higher speed of l1ls_featuresign algorithm. In addition, the average recognition rates of saliency based feature are higher than other features.

5.2 The sensitivity on the weight of sparsity term λ

For sparse representation, we apply l1ls_featuresign algorithm to compute the coefficient of sparse representation. In the sparse representation, the weight of sparsity term λ is to balance sparsity and reconstruction error, which has a certain impact on the performance of hand gesture recognition. Therefore, we conduct experiments on different values of λ respectively using the extracted gradient feature (λ =10^{-4}, 10^{-3}, 10^{-2}, 0.1, 0.2, 0.3, 0.4, 0).

From Table 2, the recognition rate is relative better when λ is in the range 0.1~0.2, so we let λ be 0.1 in our following experiments.

5.3 Performance comparison with other classifiers

To validate the effectiveness of sparse representation-based hand gesture recognition, we compare the proposed Sparse Representation (SR) method with Support Vector Machine (SVM), Artificial Neural Network (ANN) Bayesian Network (BN) and Decision Tree (DT) on saliency feature. Two hidden layers are used in ANN and the number of the nodes in hidden layers is 40 and 20 respectively. In BN method, Naïve Bayesian Network is used for recognition. In DT, we employ C4.5 algorithm for classification. The recognition results are demonstrated in Table 3, indicating that the average recognition rate of sparse representation is the highest among all the classifiers.

The success of the proposed method is partly due to the capacity of sparse representation in handling corruptions and noises, which are common in the experiment datasets. The difficulty of determining enormous parameters in ANN as well as the sensitivity to those parameters could explain the performance of ANN. While Naïve Bayesian classification method is based on the assumption that the individual components of feature vector are independent conditioned on class variable, which is not always satisfied. Furthermore, it is necessary to scan and sort dataset several times in the process of constructing tree in C4.5 algorithm, so it is less efficient.

5.4 Performance comparison on different kernel functions and different features

To validate the effectiveness of the proposed saliency based feature and kernel function, we also conduct experiments using different features and different kinds of kernel functions. Eight types of features are used including saliency, saliency histogram, surf, surf histogram, gradient, gradient histogram, lbp and lbp histogram features. Eight types of kernel functions are used as well, including Polynomial Kernel (PK): $(1+x^Ty)^b$, Inverse Distance Kernel (IDK) : $\frac{1}{1+b}\|x-y\|$, Inverse Square Distance Kernel (ISDK): $\frac{1}{1+b\|x-y\|}{}^2$, Exponential HIK (eHIK): $\sum_i \min\left(e^{bx_i}, e^{by_i}\right)$, Gaussian Kernel (GK): $\exp(-\gamma\|x_1-x_2\|^2)$, Exponential Kernel (EK): $\exp(x^Ty)$, Histogram Intersection Kernel (HIK): $\sum_i \min\left(x_i, y_i\right)$ and Linear Kernel (LK): x^Ty.

To compare performance on different features on certain kernel function, 8 types of kernel functions are all used for performance comparison of different features. Hand gesture recognition rates are shown in Figs. 2, 3, 4 and 5.

Figures 2, 3, 4 and 5 show a higher recognition rates of the proposed saliency feature than other features. The effect of λ on IDK and ISDK is larger and that on eHIK and HIK is smaller. The average recognition rates on saliency of $\lambda\in[0.1,1]$ are given in Tables 4 and 5. It is observed that the recognition rates of HIK and eHIK are highest, followed by PK and LK .

To determine the performance of various kernel functions, we use gradient, gradient histogram, lbp, lbp histogram, surf, surf histogram, saliency, saliency histogram features in experiments, the results of which are shown in Figs. 6, 7, 8 and 9. From the results, the recognition rates based on HIK and eHIK are highest and relatively not sensitive to parameter λ.

From recognition rates obtained from specific feature and different kernel functions, we can conclude that the performance of HIK and eHIK is promising and stable, which validates the effectiveness of the proposed hand gesture recognition method using saliency feature and histogram intersection kernel function.

6 Conclusions

We propose a novel saliency feature and histogram intersection kernel function based sparse representation method for hand gesture recognition. We employ l1ls_featuresign algorithm in computing the sparse coefficient. Experimental results show that the recognition rates of l1_ls algorithm are higher than those of l1ls_featuresign algorithm and sparse representation method outperforms all other classifiers compared. To sum up, saliency based feature is better than other features for all the kernel functions and histogram intersection kernel function is the best among all the kernel functions compared. Future studies include developing an effective sparse representation dictionary, new hand gesture features, and conducting view-invariant hand gesture recognition.

Acknowledgments The authors would like to thank the anonymous reviewers for the careful reading of the original manuscript. Their valuable comments and suggestions have led to a much better presentation of the paper. This research is supported in part by the Natural Science Foundation of Jiangxi Provincial Education Department under Grant No.GJJ14281 and the National Natural Science Foundation of China under Grant No. 61462038, No. 61403182, No. 61363046 and No. 61363041.

References

1. Aharon M, Elad M (2006) K-svd:an algorithm for designing over-complete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
2. Boyd SP, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge
3. Dardas NH, Georganas ND (2011) Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Trans Instrum Meas 60(11):3592–3607
4. Donoho DL, Elad M, Temlyakov VN (2006) Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans Inf Theory 52(1):6–18
5. Gao S, Tsang IW-H, Chia L-T (2013) Sparse representation with kernels. IEEE Trans Image Process 22(2): 423–434
6. Hu K, Yin L (2013) Multi-scale topological features for hand posture representation and analysis. Paper presented at the 2013 I.E. International Conference on Computer Vision
7. Just A, Rodriguez Y, Marcel S (2006) Hand posture classification and recognition using the modified census transform. In: Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on. IEEE, pp 351–356
8. Keskin C, Kırac F, Kara YE (2012) Hand pose estimation and hand shape classification using multi-layered randomized decision forests. Paper presented at the European Conference on Computer Vision
9. Kim S-J, Koh K, M. Lustig, Boyd S, Gorinevsky D (2007) An interior-point method for large-scale l1-regularized least squares. IEEE J Sel Top Signal Process 1 (4)
10. Koh K, Kim S, Boyd S (2008) l1_ls: a matlab solver for large-scale ℓ_1-regularized least squares problems
11. Krupka E, Vinnikov A, Klein B (2014) Discriminative ferns ensemble for hand pose recognition paper presented at the computer vision and pattern recognition
12. Lee H, Battle A, Raina R, Ng AY (2007) Efficient sparse codig algorithms. Advances in Neural Information Processing Systems 19
13. Li Y, Fermuller C, Aloimonos Y, Ji H (2010) Learning shift-invariant sparse representation of actions. In: Computer Vision and Pattern Recognition (CVPR), 2010 I.E. Conference on. IEEE 2630–2637
14. Poularakis S, Tsagkatakis G, Tsakalides P (2013) Sparse representations for hand gesture recognition. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Vancouver, Canada, May 26–31
15. Song Y, Demirdjian D, Davis R (2011) Tracking body and hands for gesture recognition: Natops aircraft handling signals database. FGR:500–506
16. Stergiopoulou E, Papamarkos N (2009) Hand gesture recognition using a neural network shape fitting technique. Eng Appl Artif Intell 22(8):1141–1158
17. Suk H-I, Sin B-K, Lee S-W (2010) Hand gesture recognition based on dynamic Bayesian network framework. Pattern Recogn 43(9):3059–3072
18. Tran D, Sorokin A (2008) Human activity recognition with metric learning. In: Computer Vision–ECCV 2008. Springer, 548–561
19. Wang H, Wang Q, Chen X (2012) Hand posture recognition from disparity cost map. Paper presented at the 11th Asian Conference on Computer Vision
20. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
21. Xu Z, Huang Z, Zhao Z, Li Z, Huang P (2013) Sparse representation for kinect based hand gesture recogntion system. Paper presented at the International Conference on Advanced Information and Communication Technology for Education
22. Yang X, Feng Z, Huang Z (2015) A gesture recognition algorithm using hausdorff-like distance template matching based on the main direction of gesture. Applied Mathematics and Computation 713–715:2156–2159
23. Yang W, Huang W, Duan L (2014) Multi-scale global regional contrast based salient region detection. J Comput Inf Syst 10(10):4071–4080
24. Zhang Q, Li B (2010) Discriminative k-svd for dictionary learning in face recognition. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, 2010. IEEE 2691–2698
25. Zhang L, Zhou W-D, Chang P-C, Liu J, Zhe Yan, Wang T, Li F-Z (2012) Kernel sparse representation-based classifier. IEEE Trans Signal Process 60 (4)
26. Zhou Y, Liu K, Carrillo RE, Barner KE, Kiamilev F (2013) Kernel-based sparse representation for gesture recognition. Pattern Recognition

Wenji Yang is a teacher at Jiangxi Agricultural University. She received Master's Degree from Nanchang University, China, in 2009 and Doctor's Degree at Yanshan University, China. Her research interests include image processing, pattern recognition and salient detection.

Lingfu Kong is a professor at Yanshan University. He received the Ph.D. in computer science from Harbin Institute of Technology. His research interests include computer vision, multiple visual information detection and intelligent information processing and so on.

Mingyan Wang is a professor at Nanchang University. His research interests include multiple information detection and intelligent information processing and so on.