

# Improved Collaborative Representation Classifier Based on $l_2$ -Regularized for Human Action Recognition

Shirui Huo<sup>1</sup>, Tianrui Hu<sup>2</sup>, Ce Li<sup>3,\*</sup>

<sup>1</sup>City University of Hong Kong, Hong Kong

<sup>2</sup>Beijing University of Posts and Telecommunications, Beijing

<sup>3</sup>State Key Laboratory of Coal Resources and Safe Mining, China University of Mining & Technology, Beijing

huoshirui0@gmail.com; 409346333@qq.com; \*celi@cumt.edu.cn

**Abstract.** Human action recognition is an important recent challenging task. Projecting depth images onto three Depth Motion Maps (DMMs) and extracting Deep Convolutional Neural Network (DCNN) features are discriminant descriptor features to characterize the spatio-temporal information of a specific action from a sequence of depth images. In this paper, a unified improved collaborative representation framework is proposed in which the probability that a test sample belongs to the collaborative subspace of all classes can be well defined and calculated. The Improved Collaborative Representation Classifier (ICRC) based on  $l_2$ -regularized for human action recognition is presented to maximize the likelihood that a test sample belongs to each class, then theoretical investigation into ICRC shows that it obtains a final classification by computing on which class has the largest likelihood. Coupled with the DMMs and DCNN features, experiments on depth image-based action recognition, including MSRAction3D and MSRGesture3D datasets, demonstrate that the proposed approach successfully using a distance-based representation classifier achieves superior performance over the state-of-the-arts, including SRC, CRC and SVM.

**Keywords:** Human action recognition, depth motion map features, deep convolutional neural network features, improved collaborative representation classifier,  $l_2$ -regularized

## 1. Introduction

Human action recognition has been studied in the computer vision community for decades, due to its applications in video surveillance [1], human computer interaction [2] and motion analysis [3]. Prior to the Microsoft Kinect, the conventional research focused on human action

recognition from RGB, but Kinect sensors provide an affordable technology to capture RGB and depth (D) images in real time, which can offer better geometric cues and less sensitivity to illumination changes for action recognition. In [1], a bag of 3D points and graphical model are obtained to characterize spatial and temporal information from depth images. In [3], three Depth Motion Maps (DMMs) are projected to capture body shape and motion, which is a discriminant feature to describe the spatio-temporal information of a specific action from a sequence of depth images. Seen from the literature review, although depth based methods appear to be compelling toward a practical application, even there are a few of deep-learned features for depth-based action recognition, the performance is still far from satisfactory due to the large variations of the motion. In this paper, we focus on leveraging one kind of the structure of representative model to improve performance in multi-class classification with handcrafted DMMs descriptor. In [32], three channel deep convolutional neural networks are trained to extract features of depth map sequences after projecting weighted DMMs on three orthogonal planes at several temporal scales. It was verified that the method using DCNN features can achieve almost the same state-of-the-art results on the MSRAction3D and MSRGesture3D dataset. DCNNs have been demonstrated as an effective kind of models for performing state-of-the-art results in the tasks of image recognition, segmentation, detection and retrieval. With the success of DCNN, we also take it as feature extraction and apply it in our classifier model.

As for representative models, many achievements based on space representation include image restoration [4], compressive sensing [5, 27], morphological component analysis [6], and super-resolution [7-8], *etc.* As the advances of classifiers based on representation, several pattern recognition problems in the field of computer vision can be effectively solved by sparse coding or sparse representation methods in recent decades. In particular, the advantage of linearly constrained representative models in the formulation as  $y = A\alpha$  [9], where  $y$ ,  $\alpha$  and  $A$  represent the data, a sparse vector and an given matrix with the over-complete sample set, respectively. Because of the great success of sparse coding algorithms on image processing, the sparse representation based classifiers, such as sparse representation classification (SRC) and collaborative representation classification (CRC), have gained more attention nowadays.

The basic idea of SRC/CRC is to code the test sample over a set of samples with sparsity constraints, which can be calculated by  $l_1$ -minimiztion. In [10], Wright proposed a basic SRC

model for classification by the discriminative nature of sparse representation, which is based on the theory that newly signals are recognized as the linear combinations of previously observed ones. Based on SRC, Yang *et al.* proposed a Gabor occlusion dictionary based SRC, which can significantly reduce the computational cost [11]. In [12], the authors combined sparse representation with linear pyramid matching for image classification. Rather than using the entire training set, Zhang [13] proposed learned dictionary using SRC. In [14],  $l_1$ -graph is constructed by a sparse representation subspace over the other samples. Qiao et al. [15] also proposed a method to preserve the  $l_1$ -graph for image classification by using a subspace to solve misalignment problems in image classification task. Besides, SRC is used for robust illumination [16], image-plane transformation [17] and so on. However, Zhang argued that the good performance of SRC should be largely attributed to the collaborative representation of a test sample by training samples across all classes and proposed more effective CRC. In summary, SRC/CRC simply uses the reconstruction error or residual by each class-specific subspace to determine the class label, many modified models and solution algorithms to SRC/CRC are also proposed for visual recognition tasks, including Augmented Lagrange Multiplier, Proximal Gradient, Gradient Projection, Iterative Shrinkage-Thresholding, and Homotopy [18]. Recently, some researchers [19, 20] have pointed out the purpose of  $l_1$ -regularized based sparsity in pattern classification. On the contrary, using  $l_2$ -regularized based representation for classification can do a similar job to  $l_1$ -regularized but the computational cost will reduce a lot.

Motivated by the work of modifications of CRC, in this paper, we mainly present the Improved Collaborative Representation Classifier (ICRC) based on  $l_2$ -regularized for human action recognition. Based on three DMMs descriptor feature, the ICRC approach is to jointly maximize the likelihood that a test sample belongs to each of the multiple classes, then the final classification is performed by computing on which class has the largest likelihood. The experiments on human action classification tasks, including MSRAction3D and MSRGesture3D datasets, are demonstrated and analyzed on the superior performance of this algorithm over the state-of-the-arts, including SRC, CRC and SVM. The rest of the paper is organized as follows. In Section 2, we introduce related feature descriptors using DMM. Section 3 details the action classifier based on ICRC, and Section 4 shows the experimental results of our approach on relevant datasets. The conclusion and acknowledgement are drawn

in Section 5 and Section 6.

## 2. Feature Descriptors

### 2.1 Using Depth Motion Maps

In this section, we explain the extracted feature descriptor using depth motion maps (DMMs) from depth images, which is generated by selecting and stacking motion energy of depth maps projected onto three orthogonal Cartesian planes, aligning with front( $f$ ), side( $s$ ), and top ( $t$ ) views (*i.e.*,  $p^f$ ,  $p^s$  and  $p^t$ ), respectively. As for each projected map, its motion energy is computed by thresholding the difference between consecutive maps. The binary map of motion energy provides a strong clue of the action category being performed and indicates motion regions or where movement happens in each temporal interval. We suggest that all frames should be deployed to calculate motion information instead of selecting frames. Considering both the discriminability and robustness of feature descriptors, we use the  $l_1$ -norm of the absolute difference of a frame to define the salient information on depth sequences. Because  $l_1$ -norm is invariant to the length of a depth sequence, and  $l_1$ -norm contains more salient information than other norms (*e.g.*  $l_2$ ). We have:

$$DMM^{\{f,s,t\}} = \sum_{i=1}^{N-v} |p_{i+v}^{\{f,s,t\}} - p_i^{\{f,s,t\}}| \quad (1)$$

where  $v$  is the frame interval,  $i$  represents the frame index,  $N$  is the total number of frames in a depth sequence. In the case that the sum operation in Eq. (1) is only used given a threshold satisfied, and the scale of  $v$  affects little on the local pattern histogram on the DMMs.

### 2.1 Using Deep Convolutional Neural Networks

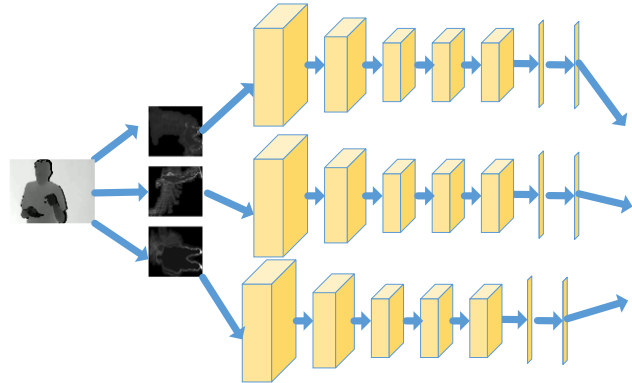


Fig.1 Three DCNNs architecture for a depth action sequence to extract features.

In this section, we introduce three deep convolutional neural networks (DCNNs) to train the features on three projected planes of DMMs, and perform fusion of three nets by combining the softmax in fully connected layer. The layer configuration of our three CNN is schematically shown in Fig. 1, in which there are five convolutional layers and three fully connected layers in each net. The detail of our implementation is illustrated in Section 4.1.2.

### 3. Action Classifier based on ICRC

Based on depth motion maps, to incorporate the feature descriptors into a powerful classifier, an Improved Collaborative Representation Classifier (ICRC) is presented for human action recognition.

#### 3.1 $l_2$ -regularized Collaborative Representation Classifier

The basic idea of SRC is to get a test sample by sparsely choosing a small number of atoms from an over-complete dictionary that contain all training samples [10]. Denoted by  $A_j \in \mathbb{R}^{d \times n}$ , the set of training samples form class  $j$ , and suppose we have  $C$  class of subjects. So  $A = [A_1, A_2, \dots, A_C] \in \mathbb{R}^{d \times n}$  involves many sample from all classes, and  $A_j$  ( $j = 1, 2, \dots, C$ ) is the individual class of training samples,  $n$  is the total number of training samples,  $d$  is the dimension of training samples. A query sample  $y \in \mathbb{R}^d$  can be presented by  $y = A\alpha$ , where  $y$ ,  $\alpha$  and  $A$  represent the data, a sparse vector and an given matrix with the over-complete training samples, respectively.

To be specific, in the mechanism of collaborative representation classifier (CRC), each data point in the collaborative subspace can be represented as a linear combination of samples in  $A$ , where  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_C]$  is an  $n \times 1$  representation vector associated with training sample and  $\alpha_j$  ( $j = 1, 2, \dots, C$ ) is the sub-vector corresponding to  $A_j$ . Generally, it is formulized as a  $l_1$ -norm minimization problem with a convex objective and solved by

$$\begin{aligned} \min_{\alpha} \{ & \|y - A\alpha\|_2^2 + \theta \|\alpha\|_1 \} \\ \text{s.t. } & y = A\alpha \end{aligned} \quad (2)$$

where  $\theta$  is a positive scalar to balance the sparsity term and the residual. The residual can be computed as

$$e_j = \|y - A_j \alpha_j\|_2, \quad (3)$$

where  $\alpha_j$  is the coefficient vector corresponding to class  $j$ . And then the output of the identity of  $y$  can be obtained by the lowest residual as

$$\text{class}(y) = \arg \min_j \{e_j\}. \quad (4)$$

The more details of SRC/CRC can be referred to [21]. Because of the computational time consuming in  $l_1$ -regularized minimization, Eq. (1) is approximated as

$$\begin{aligned} \min_{\alpha} \{ & \|y - A\alpha\|_2^2 + \lambda \|L\alpha\|_2^2 \}, \\ \text{s.t. } & y = A\alpha \end{aligned} \quad (5)$$

where  $y \in \mathbb{R}^d$  is representative by  $\alpha$  and  $A$  if the  $l_2$ -norm of  $\alpha$  is smaller. In Eq. (5),  $L$  is the Tikhonov regularization [22] to calculate the coefficient vector, and  $\lambda$  is the regularization parameter.  $\|L\alpha\|_2$  is the  $l_2$ -regularization term to add a certain amount of sparsity to  $\alpha$ , which is weaker than  $l_1$ -norm minimization. The diagonal matrix  $L$  and the coefficient vector  $\alpha$  are calculated as follows [20]:

$$L = \begin{pmatrix} \|y - h_1\|_2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \|y - h_k\|_2 \end{pmatrix}, \quad (6)$$

$$\alpha = Py \quad (7)$$

where  $P = (A^T A + \lambda L^T L)^{-1} A^T$  is independent of  $y$  and pre-calculated. With Eq. (3) and Eq. (4), the data  $y$  is assigned different identities based on  $\alpha$ .

### 3.2 The Proposed ICRC Method

Based on the training sample set, we propose an improved collaborative representation classifier based on  $l_2$ -regularized term, which assigns the data points with different probabilities based on  $\alpha$  by adding a term  $\sum_j^C \|A\alpha - A_j \alpha_j\|_2^2$  that attempts to find a point  $A_j \alpha_j$  close to the common point  $y$  inside each subspace of class  $j$ . The first two terms  $\|y - A\alpha\|_2^2 + \lambda \|L\alpha\|_2^2$  still form a  $l_2$ -regularized collaborative representation term, which encourages to find a point  $A\alpha$  close to  $y$  in the collaborative subspace. Therefore, the Eq. (5) is rewritten as

$$\begin{aligned} \min_{\alpha} \{ & \|y - A\alpha\|_2^2 + \lambda \|L\alpha\|_2^2 + \frac{\gamma}{C} \sum_j^C \|A\alpha - A_j\alpha_j\|_2^2 \} \\ \text{s.t. } & y = A\alpha \end{aligned} \quad (8)$$

Obviously, the parameters  $\lambda$  and  $\gamma$  balance three terms, which can be set from the training data. Accordingly, a new solution of representative vector  $\alpha$  is obtained from Eq. (8).

In the condition of  $\gamma=0$ , Eq. (8) will degenerate to CRC with the first two terms, and  $\|y - A\alpha\|_2^2 + \lambda \|L\alpha\|_2^2$  will play an important role in determining  $\alpha$ . When  $\gamma>0$ , these two terms  $\|y - A\alpha\|_2^2 + \lambda \|L\alpha\|_2^2$  will be same for all classes, and thus the term  $\sum_j^C \|A\alpha - A_j\alpha_j\|_2^2$  will be dominant to further fine-tune  $\alpha_j$  by  $A_j$  yielding to a precise  $\alpha$ . That is, the last newly added term is introduced to further adjust  $\alpha_j$  by  $A_j$ , resulting in a more stable solution to representative vector  $\alpha$ .

We can omit the first two same terms for all classes, make the classifier rule by the last term and formulize it as a probability exponent:

$$\text{class}(y) = \arg \max_j \left\{ \exp \left( -\|A\alpha - A_j\alpha_j\|_2^2 \right) \right\}. \quad (9)$$

The proposed  $l_2$ -regularized method for human action recognition is presented to maximize the likelihood that a test sample belongs to each class, then the experiments in following section shows that it obtains a final classification by checking which class has the maximum likelihood. So far, the above-mentioned classifier model in Eq. (8) and Eq. (9) is named as the improved collaboration representation classifier (ICRC).

## 4. Experimental Results

Based on depth motion maps, to incorporate the feature descriptors into a powerful classifier, an ICRC is presented for human action recognition. To verify the effectiveness of the proposed ICRC algorithm on action recognition applications using DMM descriptors of depth sequences, we carry out experiments on a challenging depth based action dataset MSRAction3D [1] and MSRGesture3D [1] for human action recognition.

### 4.1 Feature Descriptors

#### 4.1.1 DMMs

The MSRAction3D [1] dataset is composed of depth images sequence captured by the Microsoft Kinect-V1 camera. It includes 20 actions performed by 10 subjects facing the camera. Each subject performed each action 2 or 3 times. There are 20 action types: high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw X, draw tick, draw circle, hand clap, two hand wave, side-boxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing and pick up & throw. The size of each depth image is  $240 \times 320$  pixels. The background information has been removed in the depth data.

The MSRGesture3D [1] dataset is for continuous online human action recognition from a Kinect device. It consists of 12 gestures defined by American Sign Language (ASL). Each person performs each gesture 2 or 3 times. There are 333 depth sequences. For action recognition on the MSRAction3D and MSRGesture3D dataset, we use the feature computed from the DMMs, each depth action sequence generates three DMMs corresponding to three projection views. The DMMs of high arm wave class from the MSRAction dataset are shown in Fig. 2, and the DMMs of ASL Z class from the MSRGesture3D dataset are shown in Fig. 3.

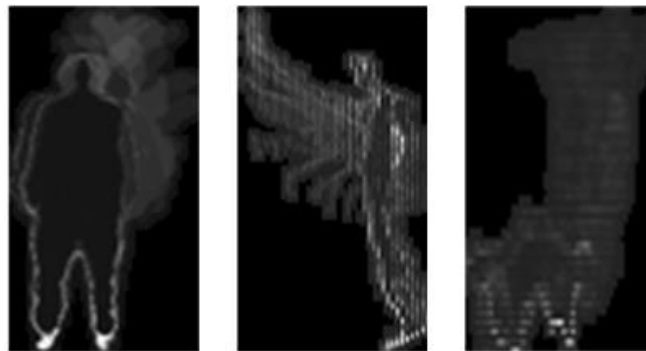


Fig.2 Three DMMs of a depth action sequence “ASL Z” from the front (f) view, side (s) view and top (t) view, respectively.

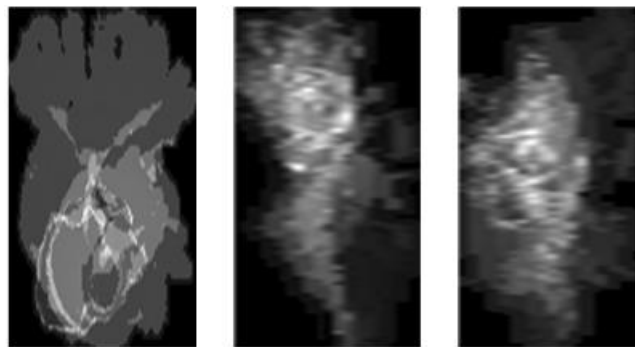




Fig.3 Three DMMs of a depth action sequence “Swipe left” from the front (f) view, side (s) view and top (t) view, respectively.

#### 4.1.2 DCNNs

Furthermore, our implementation of DCNN features is based on the publicly available MatConvNet toolbox [33] using one Nvidia Titan X card. The network weights are learnt by mini-batch stochastic gradient descent. Similar to [32], the momentum is set to 0.9 and weight decay is set to 0.0005, and all hidden weight layers use the rectification activation function. At each iteration, 256 samples in each batch are constructed and resized to  $256 \times 256$ , then  $224 \times 224$  patches are randomly cropped from the center of the selected image to artificial data augmentation. The dropout regularization ratio is 0.5 in the nets. Besides, the initial learning rate is set to 0.01 with pre-trained model on ILSVRC-2012 to fine-tune our model, and the learning rate decrease every 20 epochs. Finally, we concatenate three 4096 dimensional feature vectors in 7<sup>th</sup> fully connected layer to input the subsequent classifier.

#### 4.2 Experiment Setting

The same experimental setup in [1] was adopted, and the actions in MSRAction3D dataset were divided into three subsets as following. AS1: Horizontal wave, Hammer, Forward punch, High throw, Hand clap, Bend, Tennis serve, Pickup throw; AS2: High wave, Hand catch, Draw x, Draw tick, Draw circle, Two hand wave, Forward kick, Side boxing; AS3: High throw, Forward kick, Side kick, Jogging, Tennis swing, Tennis serve, Golf swing, Pickup throw. We performed three experiments with 2/3 training samples and 1/3 testing samples in AS1, AS2, and AS3, respectively. Thus, the performance on MSRAction3D is evaluated by the average accuracy (Accu, unit:%) on three subsets. On the other hand, the same experimental setting reported in [24-26] was followed. 12 gestures were tested by leave-one-subject-out cross-validation to evaluate the performance of proposed method.

#### 4.3 Recognition Results with DMMs and ICRC

We concatenate the sign, magnitude, and center features to form the feature based on DMMs as the final feature representation. The compared methods are similar to [24, 25]. The same parameters reported in [26] were used here for the sizes of SI and block. A total of 20 actions are employed and one half of the subjects (1, 3, 5, 7 and 9) are used for training and the remaining subjects are used for testing. The recognition performance of our method and existing approaches are listed in Table 1. It is clear that our method achieves better performance

than other competing methods.

To show the outcome of our method, Fig. 4 and Fig. 5 illustrate the recognition rates of each class in two datasets. It is stated that there are 14 classes obtaining 100% recognition rates in the MSRAction3D dataset, and the performance of 3 classes reach up to best in the MSRGesture3D dataset. All experiments are carried out using MATLAB 2016b on an Intel i7-6500U desktop with 8GB RAM, the average time of video processing gets about 26 frames per second, meeting a real-time processing demand basically.

Table 1. Recognition accuracies (unit: %) comparison on the MSRAction3D dataset and MSRGesture3D dataset.

Method	Accu. (%) on two datasets	
	MSRAction3D	MSRGesture3D
DMM-HOG [3]	85.5	89.2
Random Occupancy [28]	86.5	88.5
Actionlet Ensemble [29]	88.2	88.2
Depth Cuboid [30]	89.3	90.5
Vemulapalli [31]	89.5	87.7
DMM+CRC [26]	92.3	92.5
<b>DMM+ICRC (Proposed)</b>	<b>93.8</b>	<b>94.8</b>

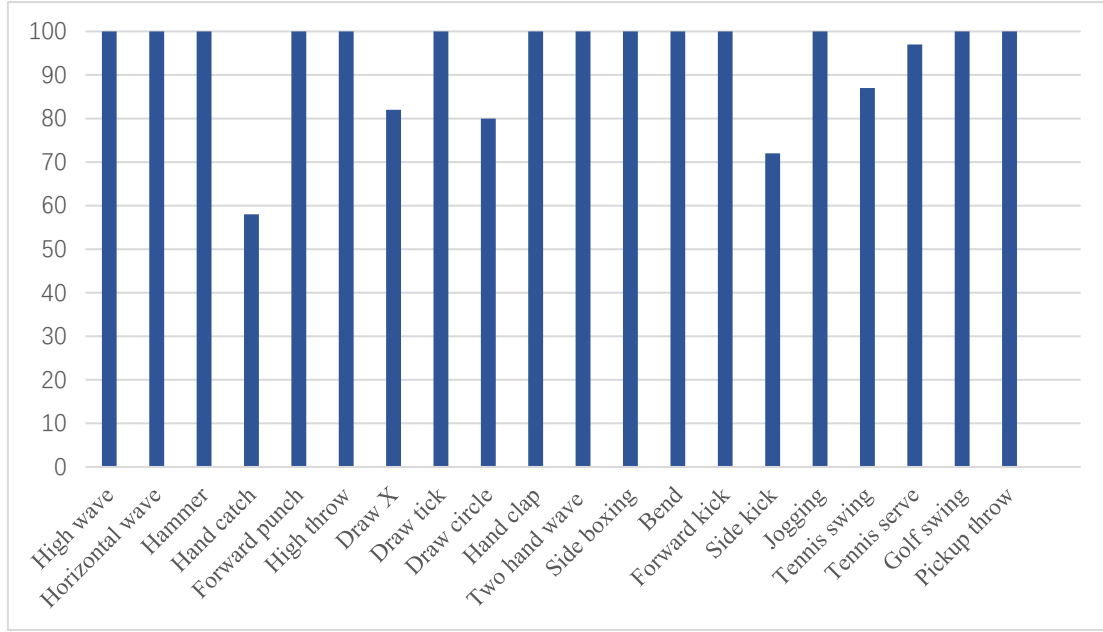


Fig. 4. Recognition rates (unit: %) of 20 class in MSRAAction3D dataset (average results of three subsets).

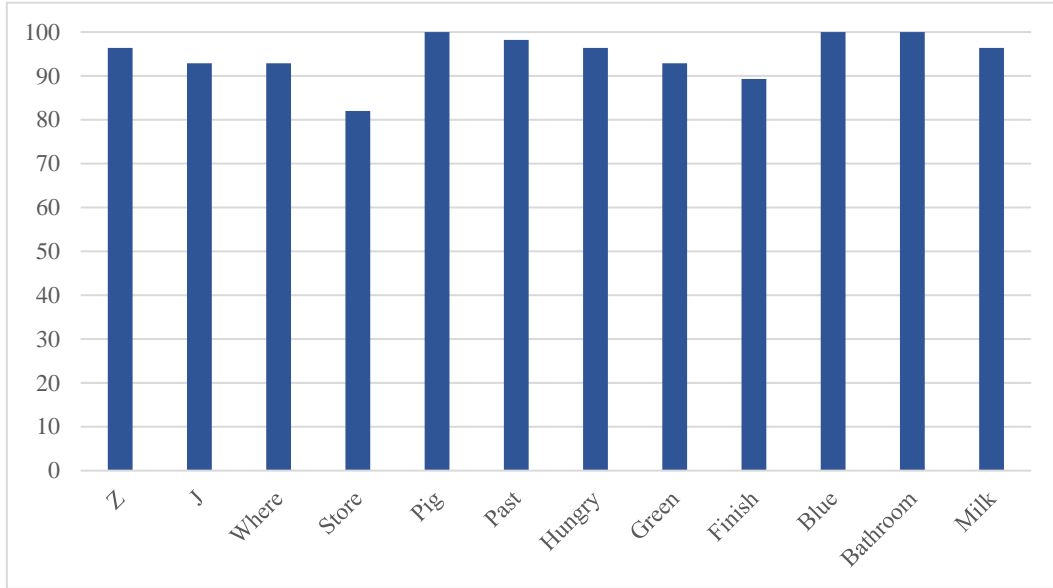


Fig. 5. Recognition rates (unit: %) of 12 class in MSRGesture3D dataset.

#### 4.4 Comparison with DCNN features and ICRC

Furthermore, in order to evaluate our proposed classifier method, we also extract the deep features by the abovementioned conventional CNN model, then input the 12288 dimensional vectors to the proposed ICRC for action recognition. Table 2 show that DCNN algorithm indeed has advances as good as in other popular tasks of image classification and object

detection, and it can improve the accuracy greatly up to 6% in MSRAction3D and MSRGesture3D. This would also explain the importance of effective feature to ICRC classifier.

Table 2. Recognition accuracies (unit: %) comparison of DMM+ICRC and DCNN+ICRC on the MSRAction3D dataset and MSRGesture3D dataset.

Method	Accu. (%) on two datasets	
	MSRAction3D	MSRGesture3D
<b>DMM+ICRC (Proposed)</b>	<b>93.8</b>	<b>94.8</b>
<b>DCNN+ICRC(Proposed)</b>	<b>99.99</b>	<b>100.0</b>

## 5. Conclusion

In this paper, we propose improved collaborative representation classifier (ICRC) based on  $l_2$ -regularized for human action recognition. The DMMs and DCNN feature descriptors are involved as an effective action representation. For the action classifier, ICRC is proposed based on collaborative representation with the additional regularization term. The new insight focuses on a subspace constraints on the solution. The experimental results on MSRAction3D and MSRGesture3D show that the proposed algorithm performs favorably against the state-of-the-art methods, including SRC, CRC and SVM. Future work will focus on involving the deep-learned network in the depth image representation and evaluating on more complex datasets such as MSR3DActivity, UTKinect-Action, NTU RGB+D, for the action recognition task.

## 6. Acknowledgments

The work was supported by the State Key Laboratory of Coal Resources and Safe Mining under Contract SKLCRSM16KFD04, in part by the Natural Science Foundation of China under Contract 61601466, in part by the Fundamental Research Funds for the Central Universities under Contract 2016QJ04. Ce Li is the corresponding author.

## Reference

- [1] W. Li, Z. Zhang, and Z. Liu, Action recognition based on a bag of 3d points, Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, pp. 9-14, 2010.
- [2] S. Wang, Z. Ma, Y. Yang, X. Li, C. Pang, and A. G. Hauptmann, Semi-supervised multiple feature analysis for action recognition, IEEE Transactions on Multimedia, vol. 16, no. 2, pp. 289-298, 2014.
- [3] X. Yang, C. Zhang, and Y. Tian, Recognizing actions using depth motion maps-based histograms of oriented gradients, pp. 1057-1060, 2012.
- [4] W.E. Vinje and J.L. Gallant, Sparse coding and decorrelation in primary visual cortex during natural vision, Science, vol. 287, no. 5456, pp. 1273-1276, 2000.
- [5] E. Candès, Compressive sampling, Proceeding of International Congress of Mathematics, vol. 3, pp. 1433-1452, 2006.
- [6] J. L. Starck, M. Elad, and D.L. Donoho, Redundant multiscale transforms and their application for morphological component analysis, Journal of Advances in Imaging and Electron Physics, vol. 132, pp. 287-348, 2004.
- [7] J.C. Yang, J. Wright, Y. Ma, and T. Huang, Image super-resolution as sparse representation of raw image patches, Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, 2008.
- [8] W. Dong, L. Zhang, G. Shi, and X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, IEEE Transaction on Image Processing, vol. 20, no. 7, pp. 1838-1857, 2011.
- [9] K. Huang, and S. Aviyente, Sparse representation for signal classification, Neural Information Processing System, 2006.
- [10] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, Robust face recognition via sparse representation, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, 2009.
- [11] M. Yang and L. Zhang, Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary, Proceeding of European Conference on Computer Vision, 2010.
- [12] J. Yang, K. Yu, Y. Gong and T. Huang, Linear spatial pyramid matching using sparse coding for image classification, Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, 2009.
- [13] Q. Zhang, and P. Li, Discriminative K-SVD for dictionary learning in face recognition,

Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2691–2698, 2010.

[14] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. Huang, Learning with l1-graph for image analysis, IEEE Transaction on Image Processing, vol. 19, no. 4, pp. 858-866, 2010.

[15] J. Yang, K. Yu, Y. Gong and T. Huang, Linear spatial pyramid matching using sparse coding for image classification, Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[16] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, Towards a practical face recognition system: robust registration and illumination by sparse representation, Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[17] J.Z. Huang, X.L. Huang, and D. Metaxas, Simultaneous image transformation and sparse representation recovery, Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[18] A. Yang, A. Ganesh, Z. Zhou, S.S. Sastry, and Y. Ma, A review of fast l1-minimization algorithms for robust face recognition, arXiv:1007.3753v2, 2010.

[19] L. Zhang, Y. Meng, F. Xiang, Sparse representation or collaborative representation: Which helps face recognition, IEEE International Conference on Computer Vision, pp. 471-478, 2011.

[20] P. Berkes, B.L. White, and J. Fiser, No evidence for active sparsification in the visual cortex, Proceeding of Neural Information Processing Systems, 2009.

[21] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210–227, 2009.

[22] A. Tikhonov, V. Arsenin, Solutions of ill-posed problems. V. H. Winston & Sons, Washington, DC, 1977.

[23] G. Golub, P.C. Hansen, D.O. Leary, Tikhonov regularization and total least squares. SIAM J Matrix Analysis Application, 21(1), 185–194, 1999.

[24] Y. Yang, B. Zhang, L. Yang, C. Chen, and W. Yang, Action recognition using completed local binary patterns and multiple-class boosting classifier, Proceedings of Asian Conference on Pattern Recognition, pp. 336–340, 2015.

[25] Campos, On the improvement of human action recognition from depth map sequences using space–time occupancy patterns, Pattern Recognition Letters, vol. 36, pp. 221-227, 2014.

- [26] C. Chen, M. Liu, B. Zhang, J. Han, J. Jiang, and H. Liu, 3d action recognition using multi-temporal depth motion maps and fisher vector, *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 3331–3337, 2016.
- [27] N. Guan, D. Tao, Z. Luo, and B. Yuan, NeNMF: An optimal gradient method for nonnegative matrix factorization, *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882-2898, 2012.
- [28] J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu, Robust 3d action recognition with random occupancy patterns, *Proceedings of the European Conference on Computer Vision*. Springer, pp. 872-885, 2012.
- [29] J. Wang, Z. Liu, Y. Wu, and J. Yuan, Mining actionlet ensemble for action recognition with depth cameras, pp. 1290-1297, 2012.
- [30] L. Xia and J. Aggarwal, Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2834-2841, 2013.
- [31] R. Vemulapalli, F. Arrate, and R. Chellappa, Human action recognition by representing 3d skeletons as points in a lie group, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 588-595, 2014.
- [32] Wang P, Li W, Gao Z, et al. Deep Convolutional Neural Networks for Action Recognition Using Depth Map Sequences. *arXiv: Computer Vision and Pattern Recognition*, 2015.
- [33] <http://www.vlfeat.org/matconvnet/>

**Article Details**

<b>Journal Title</b>	Journal of Electrical and Computer Engineering	<b>Article ID</b>	8191537
		<b>Article Type</b>	Research Article
		<b>License</b>	CC-BY 4.0
<b>Article Title</b>	Improved Collaborative Representation Classifier based on L <sub>2</sub> -regularized for Human Action Recognition	<b>Author</b> (Corresponding)	Ce Li (ORCID: 0000-0002-3081-6751)

The charges for this article are listed below.

<b>Article Processing Charges</b>	1,250.00	<b>Invoice Issue Date</b>	29 September 2017
		<b>Date of Supply</b>	29 September 2017
		<b>Invoice Ref. No.</b>	28226/2017
<b>Net Charges</b>	1,250.00		
<b>Total</b>	<b>USD 1,250.00</b>		

**Bill to**

<b>Payer Name</b>	Shirui Huo	<b>City</b>	Hong Kong
<b>Email</b>	huoshirui0@gmail.com	<b>Country</b>	Hong Kong
<b>Address</b>	the city university of Hong Kong		

The amount above is paid in full to "Hindawi Limited." This "Paid Invoice" is your receipt.

**Hindawi Limited**

Hindawi Ltd., Adam House, Third Floor, 1 Fitzroy Square, London W1T 5HF, UK  
Company Reg. No.: 08671628 VAT Reg. No.: GB 181 0941 19  
For assistance, please contact [accounts.receivable@hindawi.com](mailto:accounts.receivable@hindawi.com)