IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Journal Européen des Systèmes Automatisés
Vol. 53, No. 4, August, 2020, pp. 499-504
Journal homepage: http://iieta.org/journals/jesa

Adaptive L_p-Norm Regularized Sparse Representation for Human Activity Recognition in Coal Mines

Deyong Wang[1,2*], Zexun Geng[1]

[1] School of Information Engineering, Pingdingshan University, Pingdingshan 467000, China
[2] School of Resources and Environmental Engineering, Wuhan University of Technology, Wuhan 430070, China

Corresponding Author Email: 2548@pdsu.edu.cn

https://doi.org/10.18280/jesa.530408

Received: 8 April 2020
Accepted: 17 July 2020

Keywords:
feature extraction, sparse representation, human activity recognition, adaptive-norm regularization, structured regularization.

ABSTRACT

This paper aims to overcome the lack of in-depth exploration into the intrinsic geometry of human activities. For this purpose, a generalized adaptive L_p-norm regularized sparse representation (ARSR) approach was proposed for human activity recognition, which preserves the model adaptability through the adaptive L_p-norm regularization. In essence, the proposed method applies sparse representation to human activity recognition, turning it into a new optimization problem. In addition, the problem was solved by the iterative-shrinkage-thresholding algorithm. Specifically, the sparse representation learned by the ARSR algorithm was introduced into the support vector machine (SVM) classifier. Then, several experiments were conducted on coal-mining datasets for human activity identification. The experimental results revealed that the proposed algorithm is superior to the current sparse representation algorithms like the standard L_1-norm regularized sparse representation algorithm. The research findings shed new light on the human activity recognition in coal mines.

1. INTRODUCTION

Sequential data have become ubiquitous with the rapid development of sensors, smartphones and information technologies. For instance, human activities can be captured easily with cameras, automobile data recorders and similar devices. This attaches importance to the auto-recognition of human activities in such fields as human-computer interaction, abnormal behaviour detection, and security video surveillance. However, it is difficult to realize fast, accurate and automatic recognition of human activities due to the diverse and complex nature of such activities [1, 2].

In the field of security video surveillance, the focus of human activity recognition lies in public places like airports, railway stations, shopping malls and coal mines. The general purpose is to detect, recognize or learn interesting or dangerous events, which can be divided into such categories as suspicious, irregular, uncommon, unusual and abnormal events [3, 4]. There are four dimensions of security video surveillance, namely, tracking, recognition, motion analysis and activity analysis [5]. As the most promising research area, human activity analysis covers several topics, including gait recognition, group activity analysis and abnormality detection [6, 7]. Among them, the abnormality detection is essential to coal mining in China. Thus, the abnormality detection of coal workers is selected as the research focus in this paper.

Much research has been done on human activity recognition from both single- and hierarchical-layer approaches [8, 9]. The single-layer methods directly classify human activities from video data by representing activities as space-time features (e.g. volumes, trajectories and local features) or sequential descriptions. The typical single-layer strategies include the histograms of gradient orientations (HOG) [10], histograms of optic flows (HOF) [11] and spatiotemporal Laplacian pyramid [12]. In general, these strategies usually extract local descriptions of detected interest points, and then realize the final representation for classification using bag-of-words (BoW) or codebook quantization. The hierarchical-layer strategies recognize high-level human activities often represented by sequential concatenating and ranking of simple human actions. The typical hierarchical-layer strategies include Markov chain Monte Carlo (MCMC) [13] with Bayesian Networks [14], as well as Laplacian [12, 15] and Hessian [16] regularized approaches.

Despite the various methods, the most important issue of human activity recognition is to select a proper representation plan [17]. Inspired by the primary visual cortex (V1) modelling in the human brain, sparse representation was proposed by transforming the raw features into a new higher-level feature representation. In this approach, the conductive representations are more compact than those of the traditional BoW modelling approach [18, 19]. Recently, the regularized sparse representation approaches have been successfully applied to activity recognition, because of the efficient activity representation by regularization without losing the local structures of objects.

Nevertheless, there is a severe lack of in-depth exploration into the intrinsic geometry of human activities, such as the ignorance of the time attribute and the failure to model the learnability of sparse representation for online human activities. To solve these defects, this paper proposes a generalized adaptive L_p-norm regularized sparse representation (ARSR) approach for human activity recognition, which preserves the model adaptability through the adaptive L_p-norm regularization. In essence, the proposed method applies sparse representation to human activity

recognition, turning it into a new optimization problem.

In addition, the problem was solved by the iterative-shrinkage-thresholding algorithm. Specifically, the sparse representation learned by the ARSR algorithm was introduced into the SVM classifier. Then, several experiments were conducted on coal-mining datasets for human activity recognition. The experimental results reveal that the proposed algorithm is much better than the most recent sparsity algorithms like the standard L_1 -norm regularized sparse representation algorithm.

The remainder of this paper is organized as follows: Section 2 presents the details on the ARSR framework; Section 3 discusses the experimental results on coal-mining datasets; Section 4 wraps up this research with some meaningful conclusions.

2. ARSR

This section introduces the ARSR model and solves it with a fast iterative-shrinkage thresholding algorithm [20, 21].

2.1 ARSR framework

For N known cases $S = \{x_i\}_{i=1}^N$, S designates the sets of observed video sequence. Sparse representation aims to learn the sparse weights w_i of each example x_i simultaneously with a dictionary D. In the following, the data matrix of examples is denoted as $X = \{x_1, x_2, \cdots, x_N\} \in \Re^{m\times N}$, (here X is same as S) For a given sparse representation dictionary $D = \{D_1, D_2, \cdots, D_d\} \in \Re^{m\times d}$, the sparse representation coefficients matrix M is denoted as $W = \{w_1, w_2, \cdots, w_N\} \in \Re^{d\times N}$. Then, the L_p-norm based sparse representation [16] can be expressed as:

$$\begin{aligned} &\min_{D,W} J(D,W) = \min_{D,W} \frac{1}{2N}\|X - DW\|^2 \\ &+\lambda_1 \sum_{i=1}^{N} \|w_i\|_p \ s.t. \|D_j\|_2 \le 1, 1 \le j \le d \end{aligned} \tag{1}$$

Sparse representation under L_p-norm constraint is to find the D^*, W^* that minimize the cost function J(D, W) so that x can be approximately expressed as $X=D^*\cdot W^*$.

Under manifold assumption, the local geometry is critically important because the sparse representation w_i of example x_i and that w_j of example x_j are similar, provided that the two examples are close in the intrinsic geometry of the data distribution. Hence, the adaptive L_p-norm regularization was integrated into the cost function of Eq. (1). Thus, Eq. (1) can be rewritten as:

$$\begin{aligned} &\min_{D,W} \frac{1}{2N}\|X - DW\|^2 + \lambda_1 \sum_{i=1}^{N} \|w_i\|_p \\ &+\lambda_2 Tr(WHW^T) \ s.t. \|D_j\|_2 \le 1, 1 \le j \le d \end{aligned} \tag{2}$$

where, H stands for the HESSIAN matrix; $Tr(\cdot)$ is the trace norm.

The cost function as Eq. (2) is convex with respect to D or W, separately, but not respect to all together. In the presented algorithm, the alternating optimization was employed to optimize one of the two variables while retaining another fixed. Accordingly, the solution of the problem in Eq. (2) can be divided into two parts: sparse representation and dictionary updating, next section details this procedure.

2.2 ARSR optimization

In solving Eq. (2), two steps were involved: the first is getting sparse representation without changing dictionary, and the second is renewing dictionary while keeping the sparse representation fixed. With a known dictionary, Eq. (2) was equivalent to the follow sub-problem:

$$\begin{aligned} &\min_{W} \frac{1}{2N}\|X - DW\|^2 + \lambda_1 \sum_{i=1}^{N} \|w_i\|_p \\ &+\lambda_2 Tr(WHW^T) \end{aligned} \tag{3}$$

$$\min_{D} \frac{1}{2N}\|X - DW\|^2 \ \ s.t. \|D_j\|_2 \le 1, 1 \le j \le d \tag{4}$$

With a fixed sparse representation W, Eq. (2) can be expressed as the following:

Eqns. (3) and (4) were optimized in the following parts.

Getting Sparse Representation W with Known D. The current part optimizes Eq. (3). The general form of the sub-problem is:

$$\min_{W} F(W) = f(W) + g(W) \tag{5}$$

where, $f(W) = \min_{W} \frac{1}{2N}\|X - DW\|^2 + \lambda_1 \sum_{i=1}^{N}\|w_i\|_p + \lambda_2 \mathrm{Tr}(\mathrm{WHW}^T)$; $g(W) = \lambda_1 \sum_{i=1}^{N}\|w_i\|_p$. Both $f(W)$ and $g(W)$ are convex functions.

Eq. (5) can be converted into the following iteration plan by the gradient algorithm:

$$w_k = \arg\min_{W} \left\{ \begin{aligned} &Q_L(W, W_{k-1}) = f(W_{k-1}) + \\ &\langle W - W_{k-1}, \nabla f(W_{k-1}) \rangle \\ &+\frac{L}{2}\|W - W_{k-1}\|^2 + \lambda_1 \sum_{i=1}^{N} \|W_i\|_p \end{aligned} \right\} \tag{6}$$

In above expression, $\langle W - W_{k-1}, \nabla f(W_{k-1}) \rangle = Tr((W - W_{k=1})^T \nabla f(W_{k-1}))$; L is the Lipschitz index of ∇f. Putting aside constant term, Eq. (6) became as:

$$\begin{aligned} &W_k = p_L(W_{k-1}) = \\ &\arg\min_{W} \left\{ \begin{aligned} &\frac{L}{2}\left\| W - \left(W_{k-1} - \frac{1}{L}\nabla f(W_{k-1}) \right) \right\|^2 \\ &+\lambda_1 \sum_{i=1}^{N} \|W_i\|_p \end{aligned} \right\} \end{aligned} \tag{7}$$

From Eq. (7), it has:

$$\mathrm{w_k} = T_{\frac{\lambda_1}{L}}\left(w_{k-1} - \frac{1}{L}\nabla f(w_{k-1}) \right) \tag{8}$$

where, $T_{\frac{\lambda_1}{L}}(x)_j = max(0, |x_j|^p - \frac{\lambda_1}{L}).sgn(x_j)$.

Updating Dictionary D with Settled W. Eq. (4) is a L_2-constraint least squares question, it can be converted to:

$$\min_D \left\|X - DW\right\|^2 \tag{9}$$

This part optimizes Eq. (9) using Lagrange dual function. Taking $\beta = [\beta_1, \beta_2, \cdots, \beta_d]$ as the Lagrange multiplier, the Lagrange dual function of Eq. (9) can be written as:

$$\begin{aligned} g(\beta) &= \min_D L(D,\beta) = \min_D \left\|X - DW\right\|^2 \\ &+ \sum_{j=1}^{d} \beta_j (D_j^T D_j - 1) = \min_D Tr \\ &= ((X - DW)^T (X - DW)) \\ &+ Tr(D^T DB) - Tr(B) \\ &= \min_D Tr(X^T X - 2D^T XW^T \\ &+ W^T D^T DW + D^T DB - B) \end{aligned} \tag{10}$$

where, $B = diag(\beta)$ is a $d \times d$ diagonal matrix with diagonal elements $B_{jj} = \beta_j$ for all j.

Let the first-order derivative of $L(D,\beta)$ with respect to D be zero, we have

$$D^* WW^T - XW^T + D^* B = 0 \tag{11}$$

According to Eq. (11), D^* is obtained:

$$D^* = XW^T (WW^T + B)^{-1} \tag{12}$$

Taking Eq. (10) into Eq. (9), the Lagrange dual function of Eq. (4) can be formed as:

$$g(\beta) == \min_\beta Tr \begin{pmatrix} X^T X - XW^T \\ (WW^T + B)^{-1} WX^T - B \end{pmatrix} \tag{13}$$

Next, the Newtonian method was applied to maximize Eq. (13) with respect to β. Thus, the optimal dictionary D^* can be obtained as $D^* = XW^T (WW^T + B)^{-1}$.

2.3 ARSR algorithm

This part sums up the optimization of Eq. (3) for ARSR.

Algorithm 1 ARSR algorithm

Input: X, H, p, λ_1, λ_2
Output: W, D
Process:
Fixed D
Step 0: choose W_0, $Z_1 = W_0$, L_0, $\eta > 1$, $t_1 = 1$
Step k:
1: set $\overline{L} = L_{k-1}$
2: repeat $\overline{L} = \eta \overline{L}$, until $F(p_L(Z_k)) \le Q_L(p_L(Z_k), Z_k)$
3: set $L_k = \overline{L}$
4: update
$W_k = p_{Lk}(Z_k)$;
$t_{k+1} = \frac{1 - \sqrt{1 + 4t_k^2}}{2}$;
$Z_{k+1} = W_k + (\frac{t_k - 1}{t_k + 1})(W_k - W_{k-1})$.
Fixed W
Step 0: solve Eq. (13) to obtain β
Step 1: solve the optimal dictionary D^* by $D^* = XW^T (WW^T + B)^{-1}$
Perform iteration through the above steps until convergence.

3. EXPERIMENTS

To validate the proposed ARSR, the SVM classifier was incorporated into the sparse representation obtained by the ARSR for activity recognition. For multiple activity classes, the one-vs.-all method was employed and the binary classification was performed using an SVM classifier [22]. Several experiments were performed on a coal-mining database collected by the authors. This dataset contains 18 videos from 8 different classes of semantic activity in coal mining, including but not limited to smoking in work area, forced entry into dangerous zones and substandard operations.

3.1 Dataset and setup

The proposed ARSR method was evaluated using 18 videos containing over 20k frames. All the raw features were normalized to have zero mean so as to make the datum distribute in a symmetric interval centered on 0, and which this guarantees that the dictionary atoms towards 0 can be treated as useless members. Since the initial size of the dictionary varies with datasets, a relative small dataset was selected to reduce the computing load.

The parameters of the ARSR were configured as follows: $p = \frac{1}{2}$, and tradeoff parameters λ_1 and λ_2 were introduced to balance the regularization, the graph-based similarities and the reconstruction error. The different values of tradeoff parameters were tested on the dataset and the same parameter setting was applied to all the other scenarios. In this way, the authors prevented the heavy computing cost of the popular cross validation method.

3.2 Experimental Results

The recognition accuracy of the ARSR was contrasted with that of other spare representation –based recognition methods for human activities, namely L_2-regularization [4], Laplacian-regularization [6], Hessian-regularization [9], L_1-regularization [2] and adaptive L_1-regularization [8]. The overall recognition rate of the 8 types of activities in the dataset reached 91.88%.

Table 1. Comparison between the ARSR and the contrastive methods

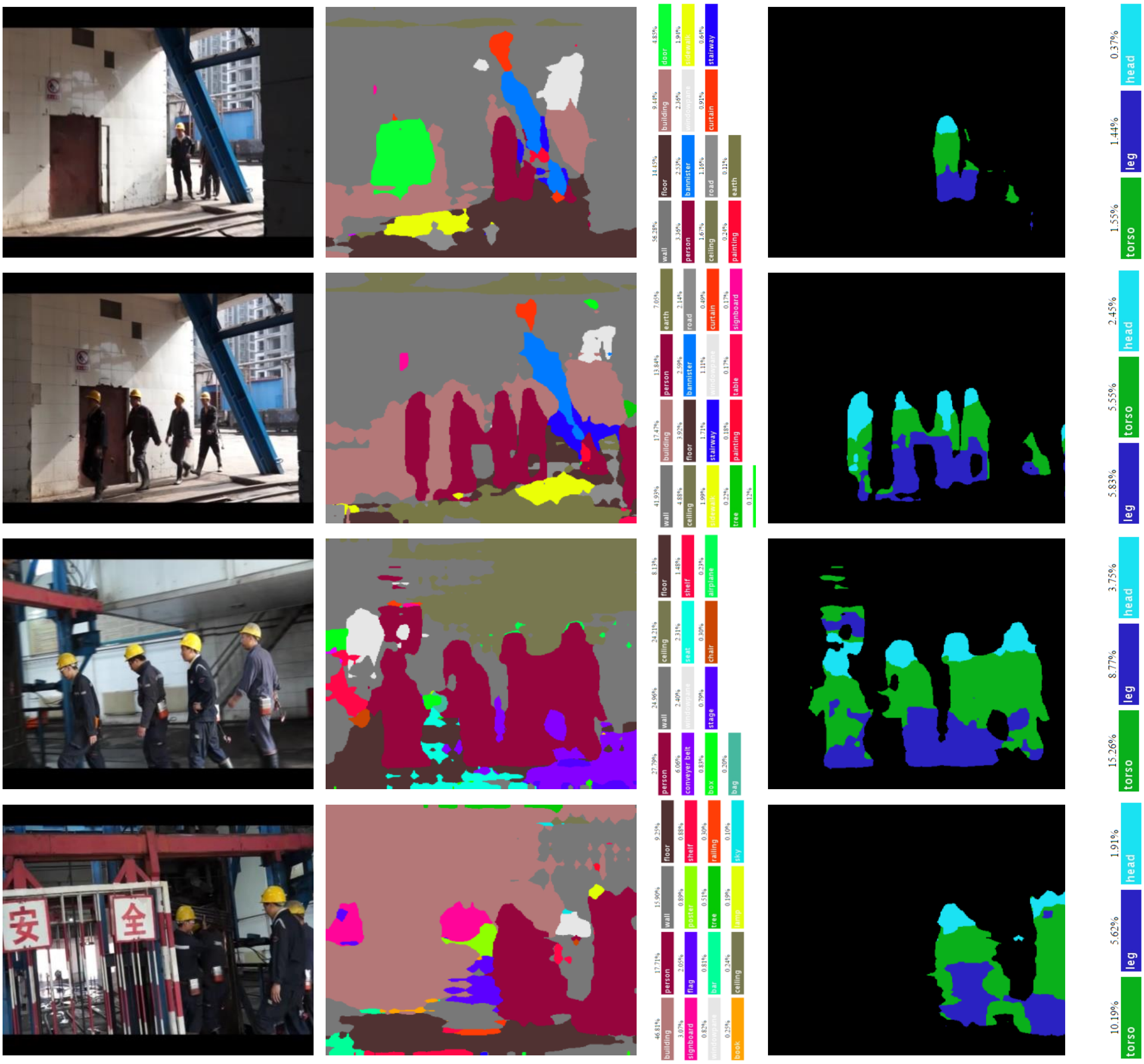
Method	L_2-regularization [4]	Laplacian-regularization [6]	Hessian-regularization [9]	L_1-regularization [2]	adaptive L_1-regularization [8]	ARSR
Result (%)	80.16	81.62	85.29	86.43	88.57	91.88

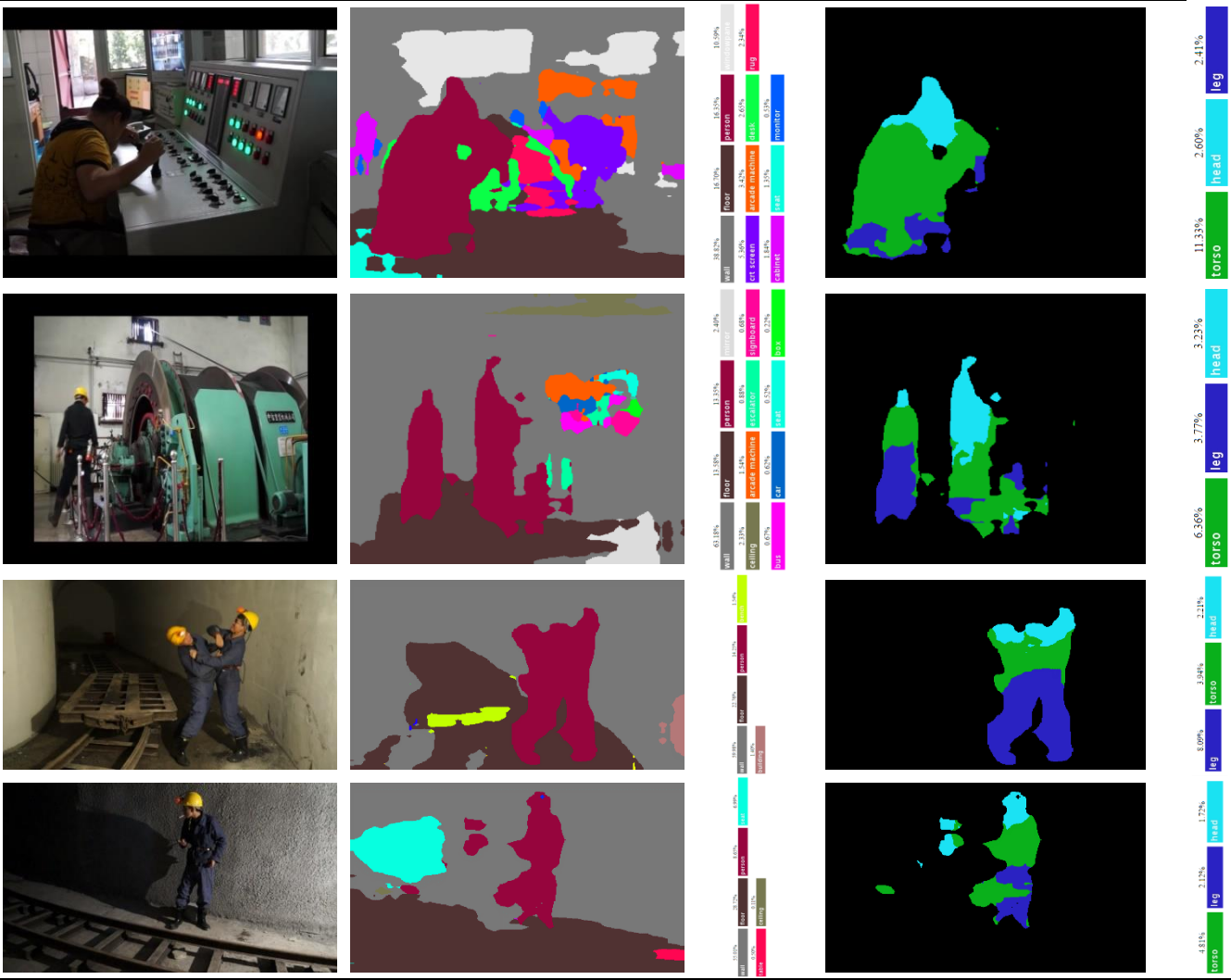
As shown in Table 1, the ARSR boasted the highest recognition accuracy. By contrast, the L_2-regularized sparse representation had the poorest accuracy. This is because the method quantifies the extracted spatiotemporal features into the dictionary without considering the inner structure of the video sequences, leading to inaccurate representation of the raw features. Compared to L_2-regularized sparse representation, both Laplacian-regularization and Hessian-regularized approaches enjoyed a relatively good accuracy thanks to the addition of a different constraint locality in the object function. The accuracy was greatly improved in both L_1-regularization and adaptive L_1-regularization. These two methods focus on the whole features in a video sequence and promote as few atoms as possible to join the encoding of all the features of the whole dataset. Hence, they achieved sparser mid-level features of every video sequence than the traditional approaches.

The proposed ARSR method absorbs the advantages of all these approaches into a unified framework and includes a fast iterative-shrinkage thresholding algorithm, leading to the best recognition accuracy. The recognition results of the ARSR are summed up in Table 2. It can be seen that the mid-level representations of behaviour learned by state-of-the-art approaches failed to separate "skipping" from "jumping forward" and "walking" of a person or many persons. On the contrary, the mid-level features learned by our ARSR can successfully tell the differences by incorporating the similarity-based structure of each activity in video sequences.

Table 2. Recognition results by the ARSR on the coal-mining dataset

Input Image	Segmentation	Objects and Their Pixel Ratios	Part Prediction	Objects and Their Pixel Ratios





4. CONCLUSIONS

Sparse representation methods have achieved promising performance in human activity recognition. The most prominent structure regularization sparse representation employs L_1-norm or L_2-norm regularization to preserve the local manifold structure. However, these methods often suffer from poor generalization. To overcome the problem, this paper proposes a generalized ARSR approach for human activity recognition, which preserves the model adaptability through the adaptive L_p-norm regularization. In essence, the proposed method applies sparse representation to human activity recognition, turning it into a new optimization problem. In addition, the problem was solved by the iterative-shrinkage-thresholding algorithm. Specifically, the sparse representation learned by the ARSR algorithm was introduced into the SVM classifier. Then, several experiments were conducted on coal-mining datasets for human activity identification, from which we could concluded that the given algorithm prevailed over the best sparse representation algorithms like the standard-norm regularized sparse representation algorithm.

ACKNOWLEDGMENT

This work was supported by the Doctor Foundation of Pingdingshan University.

REFERENCES

[1] Popoola, O.P., Wang, K. (2012). Video-based abnormal human behavior recognition—A review. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6): 865-878. https://doi.org/10.1109/TSMCC.2011.2178594

[2] Guha, T., Ward, R.K. (2011). Learning sparse representations for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(8): 1576-1588. https://doi.org/10.1109/TPAMI.2011.253

[3] Zhang, S., Yao, H., Zhou, H., Sun, X., Liu, S. (2013). Robust visual tracking based on online learning sparse representation. Neurocomputing, 100: 31-40. https://doi.org/10.1016/j.neucom.2011.11.031

[4] Qi, B., John, V., Liu, Z., Mita, S. (2016). Pedestrian detection from thermal images: A sparse representation based approach. Infrared Physics & Technology, 76: 157-167. https://doi.org/10.1016/j.infrared.2016.02.004

[5] Moayedi, F., Azimifar, Z., Boostani, R. (2015). Structured sparse representation for human action recognition. Neurocomputing, 161: 38-46. https://doi.org/10.1016/j.neucom.2014.10.089

[6] Cong, Y., Yuan, J., Liu, J. (2013). Abnormal event detection in crowded scenes using sparse representation. Pattern Recognition, 46(7): 1851-1864. https://doi.org/10.1016/j.patcog.2012.11.021

[7] Guan, Y.D., Zhu, R.F., Feng, J.Y., Du, K., Zhang, X. R. (2016). Research on algorithm of human gait recognition based on sparse representation. In Instrumentation & Measurement, Computer, Communication and Control (IMCCC), 2016 Sixth International Conference on, IEEE, pp. 405-410. https://doi.org/10.1109/IMCCC.2016.71

[8] Liu, W., Zha, Z.J., Wang, Y., Lu, K., Tao, D. (2016). p-Laplacian Regularized Sparse Coding for Human Activity Recognition. IEEE Transactions on Industrial Electronics, 63(8): 5120-5129. https://doi.org/10.1109/TIE.2016.2552147

[9] Yi, Y., Cheng, Y., Xu, C. (2017). Mining human movement evolution for complex action recognition. Expert Systems with Applications, 78: 259-272. https://doi.org/10.1016/j.eswa.2017.02.020

[10] Patel, V.M., Chellappa, R. (2013). Sparse representation-based object recognition. In Sparse Representations and Compressive Sensing for Imaging and Vision, 63-84. Springer New York. https://doi.org/10.1007/978-1-4614-6381-8_5

[11] Zhang, M., Sawchuk, A.A. (2013). Human daily activity recognition with sparse representation using wearable sensors. IEEE Journal of Biomedical and Health Informatics, 17(3): 553-560. https://doi.org/10.1109/JBHI.2013.2253613

[12] Zhang, X., Yang, H., Jiao, L.C., Yang, Y., Dong, F. (2014). Laplacian group sparse modeling of human actions. Pattern Recognition, 47(8): 2689-2701. https://doi.org/10.1016/j.patcog.2014.02.007

[13] Aggarwal, J.K., Ryoo, M.S. (2011). Human activity analysis: A review. ACM Computing Surveys (CSUR), 43(3): 1-43. https://doi.org/10.1145/1922649.1922653

[14] Theodorakopoulos, I., Kastaniotis, D., Economou, G., Fotopoulos, S. (2014). Pose-based human action recognition via sparse representation in dissimilarity space. Journal of Visual Communication and Image Representation, 25(1): 12-23. https://doi.org/10.1016/j.jvcir.2013.03.008

[15] Shao, L., Zhen, X., Tao, D., Li, X. (2013). Spatio-temporal Laplacian pyramid coding for action recognition. IEEE Transactions on Cybernetics, 44(6): 817-827. https://doi.org/10.1109/TCYB.2013.2273174

[16] Liu, W., Wang, Z., Tao, D., Yu, J. (2015). Hessian regularized sparse coding for human action recognition. In International Conference on Multimedia Modeling, Springer, Cham, pp. 502-511. https://doi.org/10.1007/978-3-319-14442-9_55

[17] Mery, D., Bowyer, K. (2014). Recognition of facial attributes using adaptive sparse representations of random patches. In European Conference on Computer Vision, Springer, Cham, 778-792. https://doi.org/10.1007/978-3-319-16181-5_59

[18] Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B. (2008). Learning realistic human actions from movies. In 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 1-8. https://doi.org/10.1109/CVPR.2008.4587756

[19] Campbell, L.W., Bobick, A.F. (1995). Recognition of human body motion using phase space constraints. In Proceedings of IEEE International Conference on Computer Vision, IEEE, pp. 624-630. https://doi.org/10.1109/ICCV.1995.466880

[20] Beck, A., Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences, 2(1): 183-202. https://doi.org/10.1137/080716542

[21] Zhu, T. (2019). New over-relaxed monotone fast iterative shrinkage-thresholding algorithm for linear inverse problems. IET Image Processing, 13(14): 2888-2896. https://doi.org/10.1049/iet-ipr.2019.0600

[22] Babu, R.V., Parate, P. (2015). Robust tracking with interest points: A sparse representation approach. Image and Vision Computing, 33: 44-56. https://doi.org/10.1016/j.imavis.2014.10.006