Review Article

Research on Human Motion Recognition Based on Data Redundancy Technology

Hong-Lan Yang,¹ Meng-Zhe Huang,² and Zheng-Qun Cai³

¹*School of Computer Engineering, Anhui Wenda University of Information Engineering, Hefei 231201, China*

²*School of Film and TV Media, Anhui Wenda University of Information Engineering, Hefei 231201, China*

³*School of Foreign Studies, Anhui Jianzhu University, Hefei 231201, China*

Correspondence should be addressed to Zheng-Qun Cai; caizhengqun1983@163.com

Received 26 February 2021; Revised 17 March 2021; Accepted 29 March 2021; Published 14 April 2021

Academic Editor: M. Irfan Uddin

Copyright © 2021 Hong-Lan Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problems of low recognition rate and slow recognition speed of traditional body action recognition methods, a human action recognition method based on data deduplication technology is proposed. Firstly, the data redundancy technology and perceptual hashing technology are combined to form an index, and the image is filtered from the structure, color, and texture features of human action image to achieve image redundancy processing. Then, the depth feature of processed image is extracted by depth motion map; finally, feature recognition is carried out by convolution neural network so as to achieve the purpose of human action recognition. The simulation results show that the proposed method can obtain the optimal recognition results and has strong robustness. At the same time, it also fully proves the importance of human motion recognition.

1. Introduction

Human motion recognition is the main research content of artificial intelligence and pattern recognition, which has attracted wide attention of scholars from all walks of life, for a long time, with the continuous progress and development of science and technology [1, 2]. Human motion recognition is faced with huge challenges due to the different influences of chaotic background and perspective changes.

Currently, humans can recognize most movements at a glance, but some remain elusive. Human movements can be set as a series of orderly poses, and there are great differences between different people or the same people when they perform the same actions under different circumstances [3]. As the types of human movements continue to increase, the overlap between each class will also increase, which brings a certain difficulty for human movement recognition.

In the following, a human motion recognition method based on data redundancy technology is proposed by combining with data redundancy technology. The experimental data analysis shows that the proposed method can realize human motion recognition quickly and accurately.

2. Methods

2.1. Image Deredundancy Processing Based on Data Deredundancy Technology. In terms of image redundancy, the wrong matching means that the correct image may be deleted, and at the same time, it will bring a series of economic losses to users [3]. In order to better deredundancy processing of the image, the following research is conducted in combination with data deredundancy technology to effectively overcome the deficiency of the deredundancy method and greatly improve the retrieval accuracy.

Through perceptual hash, images with similar structures in space are mapped to the same bucket, which can not only enhance the retrieval speed but also quickly complete the initial image filtering processing [4, 5]. Then, aiming at the images in the same bucket, the average gray feature of the block is combined with the wavelet transform to complete the image multiple filtering so as to effectively improve the image accuracy. The specific operation steps are shown in Figure 1.

Perceptual hashing is a kind of one-way mapping from multimedia dataset to perceptual digest set; that is, the digital

FIGURE 1: Image redundancy processing.

representation of multimedia with the same perceptual content is uniquely mapped to a digital digest and satisfies the perceptual robustness and security. Perceptual hashing provides secure and reliable technical support for information services such as multimedia content recognition, retrieval, and authentication. For images with similar visual perception, the hash values are similar or close to each other with high probability. However, for images with different visual perception, the probability of repeated similar hash value is significantly lower.

In order to realize image deredundancy processing, the following implementation is combined with data deredundancy technology. The specific operational steps are as follows.

2.1.1. Pretreatment. In order to enhance the speed of image preprocessing, it is necessary to process all images uniformly and make them into grayscale thumbnails with specifications of $M \times M$.

2.1.2. Extraction of Image Features. The grayscale thumbnail K is divided into N image blocks, and the average grayscale value v_i of each image block K_i and the average grayscale K of image K_{mean} are calculated, respectively, to form the block average grayscale features of image K [6]:

$$V_k = \{v_1, v_2, \dots, v_N\}. \quad (1)$$

Among them,

$$v_i = \frac{N}{M^2} \sum_{x,y \in K_i} f(x, y), \quad (2)$$

$$K_{\text{mean}} = \frac{1}{M^2} \sum_{x=1}^M \sum_{y=1}^M f(x, y). \quad (3)$$

2.1.3. Forming Hash Perception. Hash perception [7] is formed through the block average gray level feature V_k , which can be expressed in formulas (1)–(5):

$$b_i = \begin{cases} 1, & \text{if } v_i \geq v_{i+1}, i = 1, \dots, N-1, \\ 0, & \text{if } v_i < v_{i+1}, i = 1, \dots, N-1, \end{cases} \quad (4)$$

$$b_N = \begin{cases} 1, & \text{if } v_N \geq K_{\text{mean}}, \\ 0, & \text{if } v_N < K_{\text{mean}}, \end{cases} \quad (5)$$

2.1.4. Hash Clustering. Place the same image with the same perception of hashing in the same bucket.

Images with the same color sequence structure are not duplicate images; they will most likely fall into the same bucket. Thus, it can be seen that the hash perception and data redundancy technology can only accelerate the index speed of the image and complete the initial image filtering. The following needs to further filter the image repeatedly; in addition, combined with the above operation steps, it can be seen that the gray value of the average gray feature vector of the repeated image block in the same dimension is approximate, and the difference between them is low.

The following is to calculate the similarity between the two images through the Manhattan distance:

$$d(V_x, V_y) = \sum_{i=1}^N |V_{xi} - V_{yi}|. \quad (6)$$

In the above formula, V_X represents the block average gray scale feature vector of image X and V_Y represents the block average gray scale feature vector of image Y .

The block average gray feature mainly adopts the intermediate result formed by perceptual hashing filtering, so the overall operation efficiency is very fast, but the details of the point information are not fully understood. The Haar wavelet is set as the third layer of image filtering, in which the wavelet can not only accurately reflect the global information of the image but also reflect the local features of the image in detail. The specific filtering process is as follows:

- (1) Fast wavelet decomposition of the collected images;
- (2) Extract the first 60 elements with the largest absolute value from all the elements and adopt the one-dimensional form. For elements with negative absolute value, multiply their one-dimensional subscript by 1;
- (3) All the one-dimensional elements are sorted, and the sorted vector is the eigenvector;

- (4) The similarity between different feature vectors is calculated by the Manhattan distance.

To sum up, the image redundancy processing is completed.

2.2. Human Movement Recognition. For images processed by deredundancy, human action recognition is carried out by combining depth information and improved convolutional neural network, which mainly includes operation steps such as feature extraction and feature recognition. The specific operation steps are given as follows.

Depth information usually provides the pixel value of human motion image through the corresponding reference system, which is affected by different factors. In order to effectively filter out the external influence on the identification results, firstly, the depth information in the image should be preprocessed and the pixel value of the collected image should be set as the feature. In addition, in order to effectively avoid the formation of excessive dominant feature of pixels, the pixel value is normalized. The specific calculation formula is as follows:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \quad (7)$$

In the above formula, X_{norm} represents the data after normalization; X represents the initial data; $X_{\max}$ represents the maximum value of the depth action sequence; and $X_{\min}$ represents the minimum value of the depth action sequence.

Combined with the above different depth action images, feature extraction is carried out on the depth image. The specific operation steps are as follows:

- (1) Projection: Each 3D depth image frame is projected into three orthogonal Cartesian plane coordinates while each image is adjusted to the same size.
- (2) On the whole depth sequence, calculate the absolute difference of any two adjacent depth projections, and the specific calculation formula is as follows:

$$\text{DMM}_v = \sum_{i=a}^b |\text{map}_v^i - \text{map}_v^{i-1}|. \quad (8)$$

- (3) Restructuring: In order to prevent the variation of the video sequence, the size of the whole DMM is adjusted by binary and triplet interpolation.

In the process of deep learning, the input of two-dimensional convolutional neural network is usually an image, that is, a feature. If there are multiple features in the image, multiple convolutional networks are used for feature extraction and recognition, but in the whole process, the amount of calculation is too large. Using 3D convolutional network to carry out the above operations will lead to a significant increase in time and space complexity. In order to effectively solve the above problems, the convolutional neural network is comprehensively improved [7, 8]. After the improvement, multiple image features can be input at the

same time, and the convolutional kernel is two-dimensional, the overall complexity of the algorithm is also effectively reduced, and the computational efficiency is greatly improved [9].

The improved convolutional neural network is mainly composed of the following parts, respectively:

- (1) Two convolution layers: a convolution kernel is used to slide on the input matrix; data are formed according to the corresponding position of the input matrix of the convolution kernel, and all the results are added up to obtain the final convolution result. The convolution kernel is equivalent to the feature extractor, which automatically extracts the feature information of the data through the network model [10]. The calculation formula of the convolution process is given in detail in the following formula:

$$y_j = f\left(b_j + \sum_i w_{ij} * x_i\right). \quad (9)$$

- (2) Two pooling layers: the pooling layer, also known as the lower sampling layer, is equivalent to dimension reduction. It can not only reduce the amount of data but also ensure that the effective information of the image can be saved, so as to enhance the anti-distortion ability of the whole network.

Pooling can generally be divided into two different forms: average pooling and maximum pooling.

- (3) Two fully connected layers: the full connection layer is usually located behind the convolution layer and the pooling layer and usually contains one or more full connection layers in the network model. It mainly integrates local information in the convolution layer or the pooling layer to obtain the total number of feature vectors. The last full connection layer is also called the output layer, and the number of nodes in the output layer is the same as the number of categories

The improvement of convolutional neural network specifically includes the following aspects:

- (1) Improvement of convolutional neural network structure: by multidimensional input processes, two-dimensional convolution processes. Simultaneous input of multiple features can reduce the loss of image information rate and the complexity of the whole recognition process.
- (2) Addition of random deletion in the convolution layer and the full connection layer: in the training process of the whole model, random deletion will place some hidden neuron nodes in the inactive state with a certain probability, resulting in the deletion of adjacent nodes, among which these deleted nodes have no use value in the training model. However, after the above operation, the number of training parameters will be reduced, which can effectively prevent its over-fitting.

- (3) The modified linear element is used to replace the traditional saturated nonlinear function.

The specific calculation formula of the modified linear element is given as follows:

$$R_j = \max\left(0, \left(b_j + \sum_i w_{ij} * x_i\right)\right). \quad (10)$$

On the basis of the above analysis, the convolutional neural network is combined to effectively realize human action recognition.

3. Simulation Experiment

In order to verify the effectiveness of the proposed method, a simulation experiment is required. The experimental environment is Windows10 and Matlab R2017b. The experimental data come from the UCI dataset, which contains 507 training sample sets.

3.1. Comparison of Image Deredundancy Processing Results.

In the actual calculation process, the number of image blocks will have different degrees of influence on the recall rate and clustering time of the algorithm, and the experimental test is carried out on it. The human motion recognition method based on spatiotemporal image segmentation and interactive region detection proposed in literature [1] was taken as Method 1. The video human motion recognition method based on the CNN feature of the training graph proposed in literature [3] was taken as Method 2, and the specific experimental comparison results are shown in Tables 1–3.

Comprehensive analysis of the experimental data in the chart above shows that with the continuous increase in block number, the recalling rate of various methods and clustering of time in the corresponding change too, but compared with other two methods, the recall rate of the proposed method is significantly higher than that of the other two methods, and clustering time is significantly lower than the other two methods, thus fully validated the effectiveness of the proposed method and the superiority.

3.2. Recognition Rate/(%). In order to verify the accuracy of the recognition results of the proposed method, the recognition rate was selected as the evaluation index in the experiment. The higher the recognition rate, the better the recognition effect. The recognition rate comparison results of the three identification methods are given in detail Figures 2–4.

Figure 2 shows comparison results of the change of recognition rate of different recognition methods.

Based on the comprehensive analysis of the experimental data in Figure 2, it can be seen that the recognition rate of the proposed method is the highest among the three methods. The main reason is that the proposed method adopts the improved convolutional neural network for feature

TABLE 1: The image redundancy processing results of the proposed method.

Number of blocks (piece)	Recall rate (%)	Clustering time (min)
150	96.85	0.258
180	95.96	0.277
210	98.47	0.282
240	97.52	0.294
270	96.63	0.303
300	98.41	0.312
330	97.58	0.321
360	96.25	0.337
390	96.20	0.344
420	97.10	0.351

TABLE 2: Results of image redundancy processing in Method 1.

Number of blocks (pieces)	Recall rate (%)	Clustering time (min)
150	95.52	0.327
180	94.85	0.344
210	92.71	0.361
240	91.93	0.392
270	90.26	0.423
300	91.46	0.459
330	93.37	0.477
360	94.76	0.505
390	92.82	0.526
420	95.49	0.550

TABLE 3: The image redundancy processing results of Method 2.

Number of blocks (pieces)	Recall rate (%)	Clustering time (min)
150	89.85	0.450
180	87.69	0.487
210	89.14	0.535
240	88.25	0.572
270	91.96	0.603
300	93.47	0.649
330	86.19	0.671
360	81.28	0.703
390	82.37	0.724
420	84.02	0.748

recognition on the basis of the traditional recognition method, which can effectively increase the recognition rate of the whole method.

3.3. Recognition Speed (unit/second). The following experiment compares the recognition speed of the three methods, and the specific experimental comparison results are shown in Tables 4–6.

By comprehensive comparison of the experimental data in Table 6, it can be seen that the proposed method has a faster recognition speed.

FIGURE 2: Changes in the recognition rate of the proposed method.

FIGURE 3: Changes in the recognition rate of Method 1.

FIGURE 4: Changes in the recognition rate of Method 2.

TABLE 4: The recognition speed of the proposed method.

The test set	Recognition speed (PCS/SEC)
01	105
02	112
03	124
04	136
05	147
06	159
07	167
08	178
09	182
10	185

TABLE 5: Recognition speed of Method 1.

The test set	Recognition speed (PCS/SEC)
01	75
02	79
03	84
04	87
05	92
06	96
07	101
08	107
09	111
10	115

TABLE 6: Recognition speed of Method 2.

The test set	Recognition speed (PCS/SEC)
01	90
02	95
03	102
04	108
05	114
06	118
07	123
08	126
09	130
10	133

4. Conclusion

Aiming at the problems of low recognition rate existing in traditional recognition methods [11, 12], a human motion recognition method based on data redundancy technology was proposed, combining with data redundancy technology [13, 14]. Experimental results show that the proposed method can effectively eliminate the redundancy in the image, improve the recognition rate and recognition rate, and obtain a relatively satisfactory recognition effect.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare they have no conflicts of interest regarding the publication of this paper.

Authors' Contributions

All authors contributed equally to this study.

Acknowledgments

This work was supported by the Academic Project Fund for Top-Notch Talents in Disciplines (Majors) of Anhui Provincial Department of Education, no. gxbjZD2020098.

References

- [1] M. Selmi, M. A. El-Yacoubi, and B. Dorizzi, "A two-layer discriminative model for Human Activity Recognition," *IET Computer Vision*, vol. 10, no. 4, pp. 273–278, 2017.
- [2] L. Tao and R. Vidal, "Moving poselets: a discriminative and interpretable skeletal motion representation for action recognition," pp. 303–311, 2015.
- [3] S. R. Mishra, T. K. Mishra, G. Sanyal, A. Sarkar, and S. C. Satapathy, "Real time human action recognition using triggered frame extraction and a typical CNN heuristic," *Pattern Recognition Letters*, vol. 135, pp. 329–336, 2020.
- [4] Y. Du, Y. Fu, and L. Wang, "Representation learning of temporal dynamics for skeleton-based action recognition," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3010–3022, 2016.
- [5] M. Koohzadi and N. M. Charkari, "Survey on deep learning methods in human action recognition," *IET Computer Vision*, vol. 11, no. 8, pp. 623–632, 2017.
- [6] H.-J. Zhu, *Image Local Feature Detection and Description*, People's Posts and Telecommunications Press, Beijing, China, 2020.
- [7] M. Bergknut, K. Wiberg, and J. Klaminder, "Vertical and lateral redistribution of POPs in soils developed along a hydrological gradient," *Environmental Science & Technology*, vol. 45, no. 24, pp. 10378–10384, 2011.
- [8] G. Yao, T. Lei, J. Zhong et al., "Learning multi-temporal-scale deep information for action recognition," *Applied Intelligence*, vol. 49, no. 11, pp. 1–13, 2018.
- [9] X. Jiang, F. Zhong, Q. Peng, and X. Qin, "Online robust action recognition based on a hierarchical model," *The Visual Computer*, vol. 30, no. 9, pp. 1021–1033, 2014.
- [10] A. F. Liu and X. Xu, "Exploring privileged information from simple actions for complex action recognition," *Neurocomputing*, vol. 380, pp. 236–245, 2020.
- [11] H. Chen, J. Chen, R. Hu, C. Chen, and Z. Wang, "Action recognition with temporal scale-invariant deep learning framework," *China Communications*, vol. 14, no. 2, pp. 163–172, 2017.
- [12] J.-B. Liu, J. Zhao, and Z. Cai, "On the generalized adjacency, Laplacian and signless Laplacian spectra of the weighted edge corona networks," *Physica A*, vol. 540, Article ID 123073, 2020.
- [13] J.-B. Liu, J. Zhao, J. Min, and J. D. Cao, "On the hosoya index of graphs formed by a fractal graph," *Fractals-Complex Geometry Patterns and Scaling in Nature and Society*, vol. 27, no. 8, Article ID 1950135, 2019.
- [14] J.-B. Liu, J. Zhao, H. He, and Z. Shao, "Valency-based topological descriptors and structural property of the generalized sierpinski networks," *Journal of Statistical Physics*, vol. 177, no. 6, pp. 1131–1147, 2019.