July - August 2020
ISSN: 0193-4120 Page No. 5578 - 5586

Research on RGB-D Image Recognition Algorithms Based on Parallel Deep Learning

Xiaohui Zhang[1], Peng Xu [1] and Tao Xu[2,*]
[1]School of Information Engineering, Yellow River Conservancy Technical Institute, Kaifeng, Henan, 475004, China
[2]School of computer and information engineering, Henan university, Kaifeng, Henan, 475004, China

Article Info
Volume 83
Page Number: 5578 - 5586
Publication Issue:
July - August 2020

Article History
Article Received: *25 April 2020*
Revised: *29 May 2020*
Accepted: *20 June 2020*
Publication: *28 August 2020*

Abstract

Deep learning has achieved remarkable results in image recognition, using RGB-D image instead of simple feature connection. It is of great value to use RGB in depth recognition. In addition, effective feature fusion can make better use of RGB-D image instead of simple feature connection. It is of great value to use RGB in depth recognition. Based on this, this paper first describes the concept of RGB-D image recognition based on deep learning, and then analyzes the problems of the RGB-D image recognition method based on deep learning, and studies the RGB-D image recognition algorithm based on multi-mode hybrid structure, including RGB-D image recognition frame, algorithm principle, features, feature extraction and preprocessing. Finally, the RGB-D image recognition algorithm based on multi-mode hybrid structure is studied, and the experiment of image recognition algorithm is given and verified by experiment.

Keywords: *Rgb-D Image Recognition, Parallel Depth Learning, Recognition Algorit;*

1. Introduction

Deep learning is a new research hotspot in the field of machine learning. Its purpose is to build a multi-layer neural network, which can imitate the mechanism of human brain to analyze and interpret data such as images, audio and text[1-3]. Deep learning combines shallow features to form more abstract high-level features to discover deeper distributed feature representations of data[4-6]. Deep learning has caused a revolution in the field of computer vision. Many researchers and modern technology companies have focused on how to apply deep learning to various industrial fields, and have achieved some results[7-10]. At present, in-depth learning has achieved remarkable results in image recognition, scene recognition, object tracking and other aspects, showing great application value. Image recognition is one of the most important and difficult problems in the field of computer vision[11-12]. In the past research work, great progress has been made in image recognition based on RGB image and gray image. However, due to the limitations of RGB image and gray image, the application of image recognition in the field of computer vision is not very successful. For example, in the application of indoor robots, because the accuracy of recognition cannot meet the specified requirements, image recognition in the application of indoor robots once fell into a bottleneck. Improving the accuracy of image recognition is of decisive significance for the popularization of autonomous robots. The successful application of in-depth learning in the field of image recognition further promotes the development of computer vision.

Combining the advantages of deep learning in machine learning and the high efficiency of RGB-D image in object recognition, a new feature extraction algorithm is proposed to extract effective features from RGB and depth images. The RGB feature and depth feature are fused from the original image layer and decision layer using the corresponding feature fusion algorithm. Effective feature fusion can take advantage of RGB-D image better than simple feature join. Using deep learning to recognize RGB-D image has very high scientific research value and practical value.

2. Overview of RGB-D image recognition based on deep learning

Depth learning methods can be divided into supervised learning and unsupervised learning. For example, Convolutional Neural Networks (CNN) is a deep learning model under supervised learning, while DBN is the representative of unsupervised learning. In the past research work, researchers have proposed a variety of deep learning frameworks, such as deep neural network, CNN and DBN. These deep learning frameworks have been successfully applied in computer vision, speech recognition, natural language processing and other fields, and achieved good results. Object recognition, as a research hotspot in the field of computer vision and pattern recognition, has wide application value.

Image recognition is an important application direction of deep learning in the field of machine vision. The application of deep learning in image recognition has achieved some results in recent years. Traditionally, object recognition is mostly based on RGB images and gray-scale images. RGB images and gray-scale images are taken by ordinary cameras and are relatively easy to obtain. However, due to the limitations of information contained in RGB and gray-scale images, the recognition rate of complex scenes or high-resolution images is not ideal. Object recognition based on RGB image and gray image alone has gradually failed to meet the high requirement of recognition accuracy in modern industrial applications. The emergence of RGB-D cameras using a new generation of sensing technology is expected to change the status quo in the field of image recognition. RGB-D camera can take high-resolution RGB-D images, including both RGB images and depth images. RGB image contains the surface color information and texture information of the object, while depth image contains the spatial shape information of the object. It does not change with the brightness and color. RGB image and depth image are an effective complement to each other. The results show that object recognition based on RGB-D image can significantly improve the accuracy of object recognition. How to apply deep learning to RGB-D image recognition, extract body features and effectively improve the accuracy of image recognition has become a new research hotspot in the field of machine learning.

3. Current problems of RGB-D image recognition method based on deep learning

Object recognition based on RGB-D image has broad application prospects, and deep learning technology is used to complete the analysis and research of RGB-D data. In object recognition, feature set based on directional histogram design is the most commonly used feature extraction methods, such as SURF, SIFT and texture features. These methods have made some contributions to object recognition, but there are still some shortcomings. With the progress of modern industrial technology and the improvement of requirements, the features extracted by this kind of algorithm cannot meet the requirements of feature diversity and distinguishability, and cannot be well applied to new patterns, such as RGB-D images. Unsupervised feature learning algorithm is a popular method at present, which has made great progress in object recognition. However, most of the deep learning algorithms only extract features from RGB images or gray images, and have not yet been applied to three-dimensional images.

3.1. RGB-D image is difficult to apply in-depth learning framework

From the previous depth learning feature extraction algorithms are only from color image or gray image to RGB-D image. There is no distinctive feature extraction according to the characteristics of RGB-D image. That is to say, how to apply the depth learning framework to RGB-D image is still a difficult problem. The feature extraction of RGB-D image belongs to multi-modal feature extraction. We should carefully analyze and compare the multi-modal feature extraction algorithms, select the appropriate algorithm for improvement, and complete the feature extraction of RGB-D image.

3.2. Feature fusion cannot give full play to the advantages of RGB-D

In the aspect of feature fusion, depth information and color information are different and interrelated, which is a good complement to each other. In previous research work, RGB-D cannot fully play its advantages by simply connecting features in series. So it is necessary to study the latest research results of RGB-D image recognition, and learn the existing feature fusion methods, and propose a more effective algorithm to fuse RGB features and depth features.

3.3. Less feature selection and classifier selection

In feature selection and classifier selection, most of the current studies only extract color features and spatial shape features from RGB-D images, while ignoring other useful features, such as gray features and three-dimensional surface normal features. In the use of classifiers, most studies only use one classifier. For multi-modal data, the use of multi-classifiers is a very effective method to improve the recognition accuracy. Therefore, extracting more diverse features and using multi-classifier fusion technology to achieve multi-feature fusion in decision-making level has great research value in RGB-D image recognition.

4. Framework of RGB-D image recognition algorithms

The basic algorithm of RGB-D image recognition is basically consistent with the standard image recognition algorithm. The only difference is that the data input of RGB-D image recognition is multi-channel or multi-modal, i.e. color image and depth image. Because of the diversity of basic image recognition algorithms, it is unrealistic to introduce all RGB-D image recognition algorithms in detail from every perspective, because each algorithm considers not only the type of given input data, but also the way of feature extraction and classification. According to different standards, image recognition algorithms can be divided into different types. The classification criteria usually used include whether there is marked data, whether there are multiple inputs, and what form of feature extraction.

4.1. RGB-D image recognition framework

RGB-D image recognition problem is generated against the background of special data input of RGB-D image. The input in the recognition process includes not only two-dimensional color information but also three-dimensional spatial geometry information, namely four-channel input. In order to perform image recognition tasks, we need to send the given RGB-D image data into the designed image recognition framework. Although RGB-D image recognition algorithms vary in some details in different application contexts, the basic framework of most methods consists of the following four steps, as shown in Figure 1.

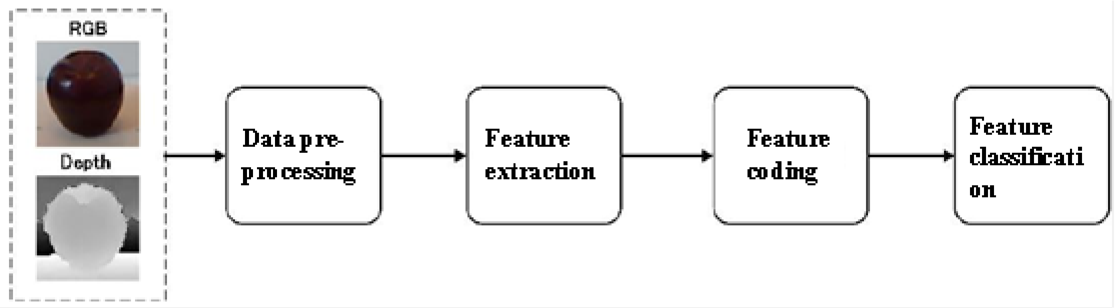


Figure1. RGB-D image recognition framework.

4.2. Principle of RGB-D image recognition algorithms

Local feature descriptor is a basic research problem in the field of computer vision. The core of this research is to construct robustness and distinguishability. Local feature descriptors play a very important role in feature description of image objects, so they are widely used in three-dimensional scene reconstruction and image recognition tasks. The commonly used image local feature descriptors are SIFT, HOG and LBP.

SIFT feature is applied to image feature extraction and image matching. It can deal with the problem of feature matching in the case of translation, rotation, scale and illumination between two images, and has the ability of resisting angle of view and affine transformation to a certain extent, as shown in Figure 2.

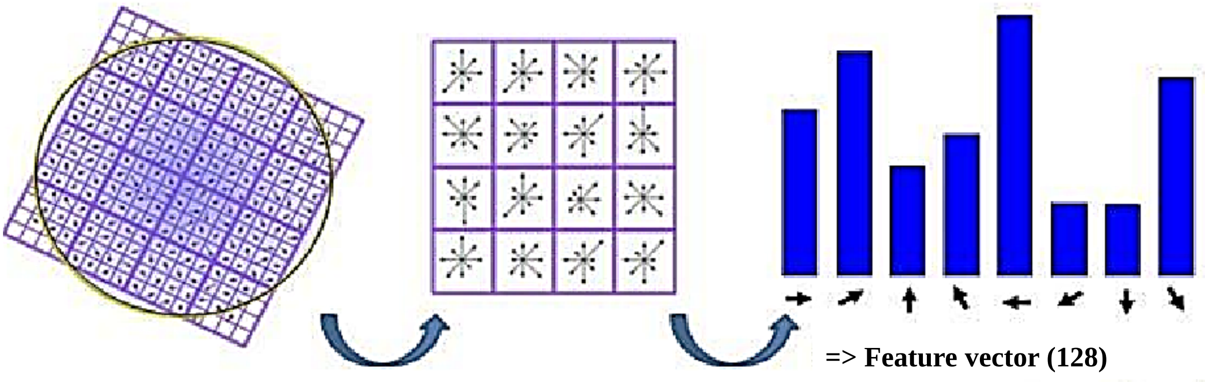
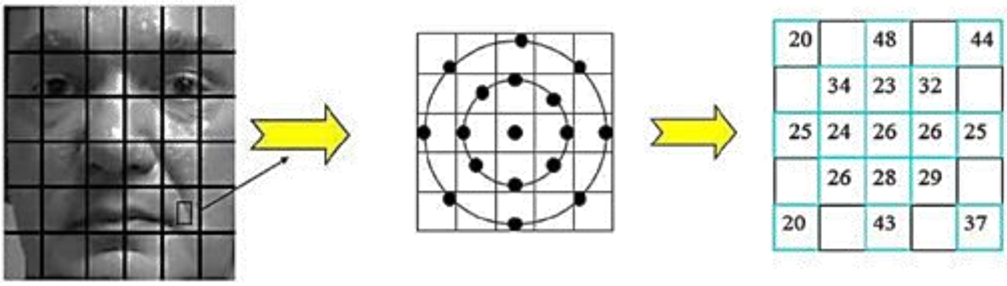


Figure2. SIFT feature description

The geometric and optical invariance of HOG feature image is essentially to calculate and statistics the gradient direction histogram of the local area of the image, which is mainly used for object detection. LBP features have rotation invariance and gray invariance. It mainly describes the local texture features of the image. The extracted features are the texture features of the local area of the image, as shown in Figure 3.



20		48		44
	34	23	32	
25	24	26	26	25
	26	28	29	
20		43		37

Figure3. LBP feature description.

4.3. Characteristics of RGB-D image

The RGB-D camera with a new generation of sensing technology captures high-resolution matched RGB images and depth images simultaneously. The pixels between RGB images and depth images correspond to each other one by one, which makes up RGB-D images. The depth image is very similar to the gray image except that each pixel value represents the actual distance between the sensor and the object. Depth image cameras are mainly used in interactive games, robots and other fields. Depth image is also called distance image. The difference between RGB image and other visible image is that from the perspective of observation, the information contained in the image is related to the distance between the object surface in the scene or an image channel.

According to the principle of depth image imaging, it can be seen that the depth image has color independence and the gray value change direction is the same as the visual direction taken by the camera. The color-independent performance of depth images does not change with illumination, shadows and environmental changes, and is not subject to too much interference. Another property, that is, the change direction of gray value is the same as the visual direction, shows that depth image can be used for 3D spatial region reconstruction tasks, and the corresponding problem of object occlusion or overlapping parts of the same object is solved, which is impossible for visible images.

4.4. Pre-processing of RGB-D image

The first step of image processing is to observe the data and obtain the characteristics of the data. Common image pre-processing algorithms include image data normalization, de-noising and whitening. Based on the research of RGB-D image, RGB image and depth image need to be pre-processed separately. The basic pre-processing can be roughly divided into three steps: image normalization, image de-noising and image whitening.

4.5. RGB-D image feature extraction based on deep learning

Combining RGB information with depth information can improve the accuracy of target recognition. Feature extraction algorithms based on deep learning include manual feature extraction and unsupervised feature learning. Manual feature extraction: Standard object recognition mainly relies on carefully designed features based on directional histogram, such as SIFT features and HOG features, and these features have achieved good results. However, these features extracted manually can only capture a small amount of recognition information. Unsupervised feature learning can learn more powerful image representation from image data. Convolutional K-means descriptor (CKM) can learn features from RGB-D images. CKM extracts features from a set of pre-detected SURF interest points. As shown in Figure 4, it can not only extract features from high resolution RGB-D images, but also combine RGB features and depth features into a more concise feature vector.

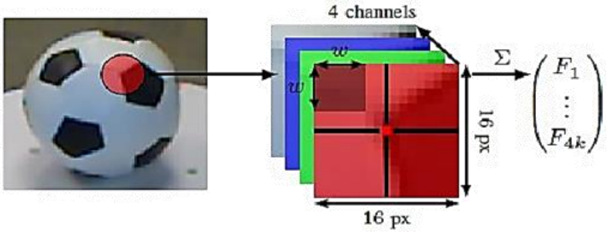


Figure 4. CKM feature learning process.

5. Research on RGB-D image recognition algorithms based on multi-modal hybrid structure

5.1. Characterization of deep mode HONV descriptor FV coding

HONV is called Histogram of Oriented Normal Vectors. The feature descriptor captures the local geometric characteristics through the angle of surface normal vectors. Usually, the surface of an object contains the most information related to the type of object. For example, we can identify the type of object by a three-dimensional mesh model without texture, but not by texture alone.

The HONV feature is shown in Figure 5, which is obtained by the azimuthal angle and zentith angle statistics of cascaded local surface regions. HONV only quantifies the spatial geometric information of depth images into the distribution of direction angles of local surface normal vectors, but it does not depict the spatial relationship between regions in depth. The improved HONV three-dimensional feature descriptor can better describe the spatial geometric information of depth images.

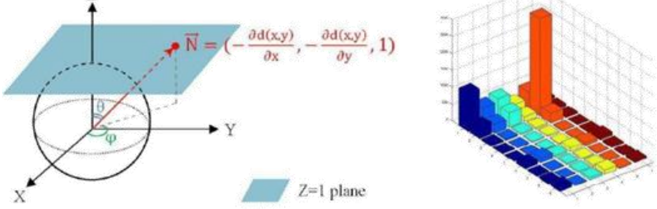

Z
$\vec{N} = (-\frac{\partial d(x,y)}{\partial x}, -\frac{\partial d(x,y)}{\partial y}, 1)$
θ
φ
Y
X
Z=1 plane

Figure 5. Normal vector direction and HONV characteristic descriptor.

5.2. Experimental results and analysis of image recognition

The open source C++ deep learning library Caffe is used to implement the deep convolution network model as a feature extractor of color modes in RGB-D images. The network responses of five network layers corresponding to VGG-16 and VGG-19 models were extracted as the characteristic expressions of RGB color modes. As shown in Figure 6, the target recognition performance of the five-level convolution response on the 10-Round test set is compared.

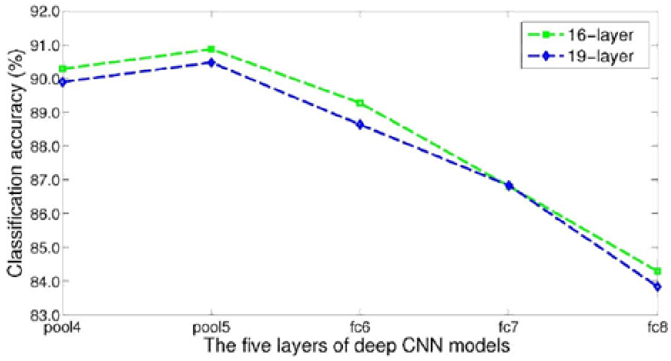


Figure6. RGB color modal recognition performance

It can be seen that the down-sampling layer of VGG-16 deep convolution neural network model has better recognition performance than other layers. Therefore, the VGG-16 deep convolution neural network model is selected as the feature extractor of the color mode of the whole RGB-D object data set, and the corresponding down-sampling layer of the network is used as the feature expression.

In the algorithm, the improved HONV three-dimensional feature descriptor is used to extract the bottom feature description from the dense sampling of depth image, and then the Fisher Vector feature coding algorithm is used to encode the middle feature expression. In order to study the recognition performance of VGG-16 deep convolution neural network model for depth image, the response of down-sampling layer is extracted as its characteristic expression for depth image. It is found that VGG-16 deep convolution neural network model has certain generalization ability, but its recognition performance is not characterized by feature description and good feature coding framework. This is because VGG-16 deep convolution neural network is trained on RGB color mode.

Compared with other RGB-D eye image recognition algorithms based on hybrid structure, as shown in Table 1, the RGB-D eye image recognition algorithm based on hybrid structure has the best recognition performance in RGB mode and depth image mode, and the overall RGB-D image recognition performance is better than other algorithms.

Table1. Comparison of RGB-D target recognition algorithms recognition performance.

Recognition methods	Depth mode	Colour mode	RGB-D
SP+HMP	82.1(±2.4)	82.4(±2.1)	86.5(±1.9)
Kernel SVM	63.5(±2.2)	74.5(±3.4)	82.9(±2.9)
Random Forest	67.8(±2.6)	73.8(±3.7)	78.6(±3.9)
Linear SVM	54.1(±1.3)	73.3(±2.3)	83.9(±2.1)
CNN-RNN	77.9(±4.8)	81.8(±3.2)	86.8(±2.3)
Hybrid structure	82.8(±1.3)	91.1(±1.2)	92.4(±1.0)

6. Experimental on RGB-D image recognition algorithms based on multi-modal hybrid structure

Multi-layer neural network algorithm based on parallel depth learning can effectively fuse RGB information and depth information in the original image layer. The algorithm is mainly divided into two layers: the first layer of deep learning network mainly completes object recognition based on RGB image and depth image separately. Through cross-validation, the recognition accuracy based on RGB image and depth image can be obtained without using test set. The decision tree algorithm is used to initialize the parameters of the MMSAE algorithm in the second layer deep learning network. The second layer of deep learning network uses MMSAE algorithm to extract features from RGB images and depth images simultaneously, and completes the fusion of RGB features and depth features. Then, it uses spatial pyramid maximum pooling algorithm to extract more abstract and distinctive high-level features. Finally, the final recognition accuracy is obtained by using Softmax classifier.

6.1. Image recognition based on RGB and depth information

For different images, the recognition effect using RGB and depth features is different. The higher the recognition accuracy based on RGB features is, the more effective the recognition of RGB features is for such images, and the greater the contribution rate of recognition is. On the contrary, the same is true for depth features. In order to get the difference of recognition effect, the first layer of deep learning network uses RGB image and depth image to recognize the image, and constructs decision tree by the recognition accuracy, which initializes the parameters of the second layer network's MMSAE algorithm. Decision tree algorithm is the initialization parameter of MMSAE algorithm: In machine learning, decision tree is a prediction model, which maps object attributes to object values. Figure 7 is a schematic diagram of the decision tree used.

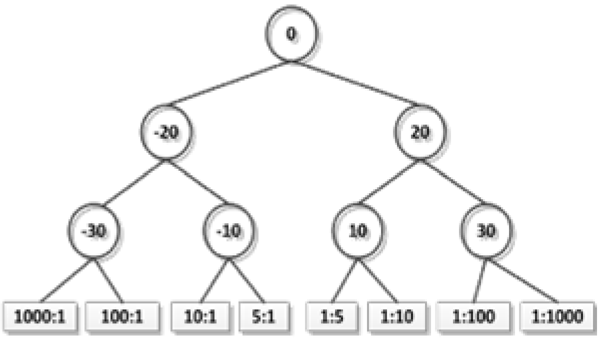


Figure 7. Structure of decision tree algorithm.

6.2. Raw image layer fusion of RGB features and depth features

After obtaining the recognition accuracy of RGB image and depth image, the decision tree algorithm is used to initialize the parameters of MMSAE algorithm. This parameter will play a decisive role in MMSAE algorithm. Appropriate parameters will effectively extract more effective features from RGB images and depth images, thus improving the recognition accuracy. The second layer of the network can be divided into the following steps: Firstly, MMSAE algorithm extracts shallow features. Secondly, the spatial pyramid maximum pooling algorithm extracts high-level Abstract features. Finally, the final recognition accuracy is obtained by using the Softmax classifier: after the final features are obtained, the classifier is trained by using the training set, and the final recognition accuracy is obtained by using the test set.

6.3. Experimental verification

Using the 2D 3D database, which contains common objects, each object contains 40 pairs of RGB images and depth images, which are taken by placing the object on the tray at 30 degrees of rotation. In each experiment, the database is randomly divided into two parts, one as a training sample and the other as a test sample. The segmentation rule is to randomly select five objects from each category for training and the remaining objects for testing. For samples with less than 5 objects in each category, one object is randomly selected for testing, and the remaining objects are used for training to ensure that at least one object in each category is used for testing. For each object, 18 images with uniform angular distribution are selected for training or testing. The whole experiment was repeated 40 times, each time randomly partitioning the database, and the final experimental results were taken as the average of 40 experimental results.

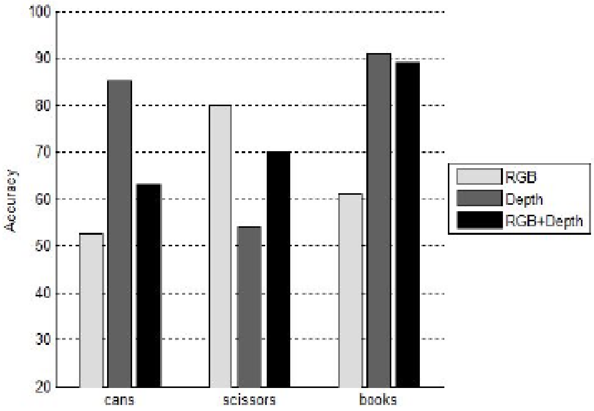
The experimental results are shown in Table 2. This algorithm effectively completes the fusion of

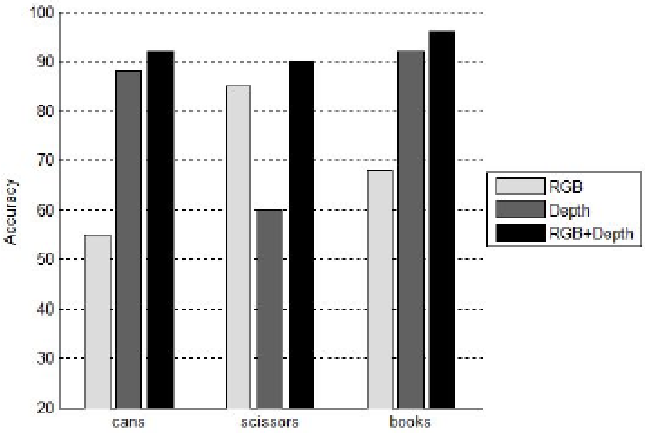
RGB information and depth information, and improves the recognition accuracy of RGB-D image. It shows that the MMSAE algorithm can effectively extract and fuse the differences between RGB features and depth features, and can effectively extract and fuse the shallow distinguishing features from multi-modal data. The new depth model completes the feature extraction and recognition of RGB-D image well, and improves the recognition accuracy of RGB-D image. At the same time, the experimental results show that the effective fusion of RGB features and depth features can give full play to the advantages of RGB-D image better than the simple linear connection between RGB features and depth features.

Table2. Comparing results of database experiments.

Algorithm	RGB + Depth
SP+HMP	91.3
ICCV workshop	82.8
Experimental algorithm	94.1

In the acquisition process of 2D 3D database, the recognition accuracy of some objects based on RGB features or depth features is relatively low due to the low discrimination of RGB information or depth information of some objects or the lack of information. Based on this, the RGB feature and depth feature are connected in series and used as the final feature of the object for object recognition, which will not improve the accuracy of object recognition, but will lead to the final recognition accuracy lower than that based on the RGB feature or depth feature alone, or even lower than both.





(a) (b)

Figure 8. Recognition accuracy of object features.

From Figure 8(a), it can be seen that for some images, the recognition accuracy based on parallel depth features is significantly higher than that based on RGB features. After linking the two features, the recognition accuracy is lower than that based on the depth feature alone. For some objects, the recognition accuracy based on RGB features is significantly higher than that based on depth features. After linking the two features, the recognition accuracy is lower than that based on RGB features alone. Therefore, inappropriate feature links will not give full play to the advantages of RGB-D images.

Figure 8 (b) Relative to the recognition accuracy of the corresponding image, clearly reflects the effective feature fusion, image recognition accuracy based on fusion features is significantly higher than that based on RGB features and depth features alone. The inappropriate feature fusion method can eliminate the situation that the recognition accuracy based on fused features is lower than that based on individual features.

7. Conclusion

Image recognition has important research value and application prospects. It is also a research hotspot in aerospace, meteorological observation, automated agriculture, medical image analysis and other fields. RGB-D target recognition algorithm based on hybrid structure has achieved excellent recognition performance on large-scale RGB-D object data sets. The overall recognition performance of RGB-D exceeds that of other best algorithms. Aiming at the task of RGB-D stereo target recognition, the RGB-D image recognition algorithms based on parallel in-depth learning are proposed, and the performance of the proposed algorithm is analyzed and evaluated on large data sets. In the aspect of fusion of RGB image and depth image features in decision-making layer, firstly, more kinds and more diverse features are extracted from RGB-D image. Combining with multi-classifier system, the fusion of multiple features in decision-making layer is completed by static linear combination. The experimental results show that the fusion of multiple features at the decision level is also an effective way to make full use of the complementary advantages between RGB images and depth images.

References

[1] Xu ke. Research on the application of convolutional neural network in image recognition [D]. Zhejiang University, 2012.

[2] Fox D, Bo L, Ren X. RGB-(D) scene labeling: features and algorithms[c]// computer vision and pattern recognition. IEEE, 2012:2759-2766.

[3] Ren X, Fox D. Unsupervised feature learning for RGB-D based object recognition[C].Experimental Robotics. Springer International Publishing, 2013: 387-402.

[4] Hinton G. A practical guide to training restricted boltzmann machines[J]. Momentum, 2010, 9(1): 926.

[5] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]. Proceedings of the ACM International Conference on Multimedia. ACM, 2014: 675-678.

[6] Xu C, Tao D, Xu C. A survey on multi-view learning [J]. arXiv preprint arXiv:1304.5634, 2013.

[7] Perronnin F, Mensink T, et al. Image classification with the fisher vector: Theory and practice [J]. International Journal of Computer Vision, 2013, 105(3): 222-245.

[8] J. Yang, B. Price, S. Cohen, H. Lee, and M. H. Yang. Object contour detection with a fully convolutional encoder-decoder network. In CVPR, 2016.

[9] Jia Lei, Chen Yuqiang, et al. Yesterday, today and tomorrow of in-depth study [J]. Computer Research and Development, 2013, 50 (9): 1799-1804.

[10] Liu Ran. Binocular Stereo imaging research based on computer stereo vision [D]. Chongqing: Chongqing University, 2007.

[11] Hansard M, Lee S, Choi O, et al. Time-of-flight cameras: principles, methods and applications[M]. Springer Science & Business Media, 2012.

[12] Zhang Z. Microsoft Kinect Sensor and Its Effect[J]. IEEE Multimedia, 2012,19(2):4-10.