Summary of Object Recognition

Shiji Li1a, Hong Chen1*, Qing Wang1, Jiahui An1 and Jiayue Li1

1Department of Computer Science and Technology, China Agricultural University, Beijing, 100083, China

a sy20193081482@cau.edu.cn  * 597270947@qq.com

Abstract. Object recognition is one of the classic problems in computer vision. It is very important for computers to recognize the common objects in life like the human brain and the human eye. This is also an important step in the development of computers in the direction of intelligence. This article summarizes the current research on object recognition, and elaborates on several aspects of object feature extraction, object feature matching and object recognition methods.

1. Introduction

Object recognition technology is a basic research in the field of computer vision. Its main function is to recognize what object is in the image and give the position and direction of the object in the image, so that the computer can simulate the human brain and human eye function. This technology can be applied in many fields, such as face recognition system, augmented reality, virtual reality, machine vision and other fields.

In recent years, the development of computer hardware and software and the introduction of efficient recognition and detection algorithms have led to the development of object recognition technology towards the four criteria for evaluating object recognition methods, namely robustness, correctness, efficiency and scope. The process of object recognition is divided into the processes of obtaining a single frame image of an object, image preprocessing, feature extraction, feature selection, feature matching, and object identification information feedback from the matching result. The key to whether an object recognition technology is efficient lies in the efficiency of the object feature extraction, feature processing matching, and classification and recognition methods. With the continuous development of deep learning, the application of deep learning in the field of object recognition has become a research hotspot in various enterprises and scientific research institutions.

Current object recognition technologies can be classified into two categories: model-based or context-based recognition methods; two-dimensional object recognition or three-dimensional object recognition methods.

This article summarizes the current research on object recognition, and introduces in detail the aspects of feature extraction, feature matching, and recognition methods, and finally proposes to summarize the current situation and future prospects.

2. Research on object feature extraction

Computer recognition of objects is the same as the process of human beings at the beginning of recognizing objects. First, they must recognize the characteristics of the object, and this characteristic is the feature of the object. Therefore, feature extraction is the first step in the main process of object recognition and an important part of the recognition method. Good object features make objects have
better separation in the high-dimensional space of the image, which can reduce the burden of the follow-up process of the recognition algorithm. The following introduces commonly used feature extraction methods from the perspectives of color features, texture features, and shape features.

Color features are visual information often used in image recognition of objects. Compared with other features, color features are less affected by image size, direction, and viewing angle. The use of color features should be based on different color spaces, the color spaces mainly include RGB color space, HSV color space, HIS color space, etc. The HSV color space is the most frequently used space, because it defines colors based on humans’ intuitive perception of color, light and shade, and tone, and its three components represent hue, saturation and value, which is in line with human intuitive perception. Commonly used color description methods include color histograms, color moments, color sets, color aggregation vectors, etc. Among them, color histograms have become the most commonly used color feature description methods due to their good robustness [1]. Xue Xuqin [2] proposed a deep residual network to extract the color features of vehicles through autonomous learning, and add color space information to improve the accuracy of vehicle color recognition. Li et al. [3] extracted the color histogram by quantizing the H and S components equally in the HIS color space, and normalized the color histogram features to perform template matching and complete vehicle color recognition.

Texture features is the local nature of the image. It is the pattern obtained by transforming the grayscale in some form in space [4], which reflects the relationship between the image surface information and its surrounding environment [5]. In general, texture features can provide sparseness, smoothness, regularity and other features of the image area. Almost all images contain texture information. Texture features extraction methods are mainly divided into statistical method, frequency spectrum method and structure method. The most commonly used statistical methods are gray level co-occurrence matrix (GLCM), local binary pattern (LBP) and Tamura texture feature description method. Frequency spectrum method mainly uses wavelet transform to extract image features. Structure method is seldom applied in actual research and has gradually faded out of everyone's vision. Texture features extraction methods can be combined to improve the efficiency of object image recognition and retrieval. Ren et al. [6] proposed a new algorithm combining GLCM and discrete wavelet decomposition. Zhang Lei [7] proposed a new algorithm for image LBP-GLCM feature extraction.

Shape features extraction is a quantitative and qualitative analysis of the shape of an object. In order to recognize objects, researchers must describe the shape of the object. The shape features of the same object should remain unchanged in different positions, sizes and directions of the image. The description methods of shape features are generally divided into two categories: contour-based and region-based [8]. There are three main methods for describing shape features based on boundary contours: Fourier descriptors, wavelet descriptors, and boundary histograms [9]. There are three methods for region-based shape features description: Hu invariant moments, generalized Fourier descriptors and Zernike moments. Ling et al. [10] improved the shape context method by calculating the inner distance between the contour points through the internal connections of the shape instead of the Euclidean distance, and established the inner distance shape context histogram.

In addition, the use of neural networks to extract object image features is also a commonly used feature extraction method. Its essence is to replace manually defined feature image extractors and manually defined models to automatically learn and extract features. In recent years, the more commonly used neural network object detection and recognition methods mainly include YOLO, RCNN, SSD and RetinaNet.

3. Research on object feature matching
In the research of feature matching technology, the feature similarity measurement methods between different object images can be divided into two categories: set theory model and geometric matrix model. The commonly used method is the geometric matrix model. Its main idea is to treat the eigenvectors of two images as two points in the feature space, and calculate the distance between the two points to measure the similarity between the two images. The following are several commonly used similarity
measurement methods, where x and y represent the eigenvectors of two images, and their i-th component is represented as \( x_i \) and \( y_i \).

Minkowsky distance \(^{[11]}\). The McCausky distance is a summary of several distance measurement formulas, expressing a variety of distance measurement formulas. The expression is as follows:

\[
D(x,y) = \left( \sum_{i=1}^{m} |x_i - y_i|^p \right)^{1/p}
\]

Euclidean distance \(^{[12]}\). The size of the Euclidean distance indicates how similar two samples are located in the area. The larger the Euclidean distance, the more similar the region where the sample is, and the higher the matching degree; the smaller the Euclidean distance, the less similar the region where the sample is, and the lower the matching degree. The expression is as follows:

\[
D(x,y) = \left( \sum_{i=1}^{m} (x_i - y_i)^2 \right)^{1/2}
\]

Manhattan distance. The expression is as follows:

\[
D(x,y) = \sum_{i=1}^{m} |x_i - y_i|
\]

Canberra distance. The expression is as follows:

\[
D(x,y) = \sum_{i=1}^{m} \frac{|x_i - y_i|}{x_i + y_i}
\]

Chebychev distance. The expression is as follows:

\[
D(x,y) = \max_{1 \leq i \leq m} |x_i - y_i|
\]

Although scholars have proposed many distance measurement formulas, the most commonly used formula is Euclidean distance.

4. Research on object recognition method
Object recognition methods are roughly divided into two categories: model-based or context-based recognition methods and two-dimensional feature-based or three-dimensional feature-based recognition methods

4.1. Model-based object recognition method
The objects processed by this method are object images and object models. The feature matching algorithm needs to find the most similar one from the object model to the real object image. Its main job is to find the correspondence between the features extracted from the two-dimensional or three-dimensional image and the built-in features in the model library to predict what the object is.

4.2. Context-based object recognition method
In the model-based object recognition method, the scene where the object is located has a negative impact on the recognition of the object. However, in the real world, the scene where the object is located provides a lot of useful information for people to recognize the object. This is because the object generally does not appear in a certain scene out of thin air, and its appearance is always related to the scene. The context-based object recognition method is to use the scene information of the object to better recognize and interpret the object. Lee et al.\(^{[13]}\) used experiments to verify that objects can be recognized more effectively when the context and appearance of objects are used.

4.3. Two-dimensional feature-based object recognition method
The object processed by this method is a two-dimensional image of an object. The two-dimensional feature mentioned here mainly refers to color features, texture features, and shape features of the object in the two-dimensional image. This method first extracts the features of the object image, and takes
appropriate processing to the features to suit the recognition of such objects. Two methods are mainly used in feature matching. One is to calculate the distance relationship with the feature of the reference image in the database to predict what the object is; the other method is to use a classifier to classify the object image. Commonly used classifiers are Support Vector Machine (SVM), Adaboost, Boosting, Decision Tree.

4.4. Three-dimensional feature-based object recognition method

Object recognition based on three-dimensional features is an object recognition direction with a large research space at present. The difference between it and the general object recognition problem is mainly reflected in the known conditions (often the three-dimensional model of the known object), or the detected and recognized target (the three-dimensional bounding box, and the approximate posture of the object). Generally speaking, some three-dimensional features of the object must be known. Compared with two-dimensional features, the acquisition of three-dimensional features requires measuring the object, and then calculating the corresponding three-dimensional data through some geometric principle formulas, or knowing the three-dimensional model of the object in advance. The degree of conditional acquisition is more difficult. In some cases, the algorithm effect is better than the two-dimensional features algorithm. This type of algorithm generally uses the three-dimensional point cloud data of the object as the input for detection and recognition. Martin et al.[14] used a three-dimensional sliding window method, first rasterizing the point cloud, then using a fixed-size three-dimensional window, and using CNN to determine whether the window area is a vehicle.

5. Conclusion

Object recognition technology has experienced decades of development. From traditional computer vision technology to the current popular deep learning recognition technology, the recognition effect is more and more accurate, and the algorithm is more robust. But the better the effect and efficiency of the algorithm means the improvement of the network and hardware requirements, which is a relatively large restriction for the application of the algorithm. And the current recognition algorithm based on deep learning requires a sufficient number of training sets to train the model to achieve a good recognition effect. This is not a priority for objects without a large number of similar object data sets, such as some rare collections in museums. In addition, the current object recognition almost always requires network transmission of images, and then transmission of recognition feedback results through the network, which is much less efficient in environments with relatively poor network conditions, such as the wild and crowded indoor environments.

In summary, the future development prospects of object recognition technology should not be limited to the recognition accuracy, recognition efficiency, algorithm robustness, etc., but also specific issues should be considered. For objects with different characteristics and the environment in which the objects are located, designing corresponding recognition algorithms and considering the degree of dependence on network and equipment performance should also be a focus of future research.

Acknowledgments

Thanks to my mentor Chen Hong and Wang Qing. Thanks for their valuable suggestions for my article.

References

[1] Tsang, W.H, Kosor, A, Kwok, T.Y.(2008)Large-Scale Maximum Margin Discriminant Analysis Using Core Vector Machines. IEEE Transactions on Neural Networks,19(4):610-24.
[2] Xue, X.Q.(2018)Research Vehicle Color Recognition Scheme Based on Machine Learning. Beijing University of Posts and Telecommunications.
[3] Li, X.Z, Zhang, G.M, Fang, J, et al.(2010)Vehicle Color Recognition Using Vector Matching of Template. In Third International Symposium on Electronic Commerce and Security . IEEE Computer Society.189-193.
[4] Jiang, J.R.(2016)Research and Implementation of Network Texture Automatic Acquisition Algorithm Based on Synthesizable Measurement. Shenzhen University.

[5] Khalvati, F, Wong, A, Haider, M.A.(2015)Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. Bmc Medical Imaging,15(1):27.

[6] Ren, L.H, Zhang, H.Y, Guo, K.M.(2012)Color texture recognition methods base on GLCM-Gabor and application in print detection. International Journal of Advancements in Computing Technology.

[7] Zhang, L.(2015)Fabric defect classification based on LBP and GLCM. Journal of Fiber Bioengineering & Informatics,8(1): 81-89.

[8] Zhang, D.S, Lu, G.J.(2004)Review of shape representation and description techniques. Pattern Recognition,37(1-19).

[9] Huang, W.G, Gu, C, Zhu, Z.K.(2013) PCA-SC shape matching for object recognition. Optics & Precision Engineering,21(8):2103-2110.

[10] Ling, H.B, Jacobs, D.W.(2007)Shape Classification Using the Inner-Distance. IEEE Transactions on Pattern Analysis and Machine Intelligence,29(2):286-299.

[11] Dunne, P, Matuszewski, B.(2011)Choice of similarity measure, likelihood function and parameters for histogram based particle filter tracking in CCTV grey scale video. Image and Vision Computing,29(2):178-189.

[12] Lee, S.M, Xin, J.H, Westland, S.(2005) Evaluation of image similarity by histogram intersection. Color Research and Application,30(4):265-274.

[13] Lee, J.Y, Grauman, K.(2010)Object-Graphs for Context-Aware Category Discovery. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition,1-8.

[14] Engelcke, M, Rao, D, Wang, D.Z, et al.(2017)Vote3Deep: Fast Object Detection in 3D Point Clouds Using Efficient Convolutional Neural Networks. IEEE International Conference on Robotics and Automation (ICRA),1355-1361.