DOI: 10.7763/IPEDR. 2012. V49. 18

The Intervention of Artificial Intelligence in Image Semantics Retrieval

Yujiao Wang and Haiyun Lin

Department of Physical Science and Technology, Kunming University, Kunming, China

tjwyj817@163.com, lhy198@163.com

Abstract—In respect of the classification of current image retrieval technology and the existing issues, the paper put forward a method designed for image semantic feature extraction based on artificial intelligence. The new method has solved the tough problem of image semantic feature extraction, by fusing fuzzy logic, genetic algorithm and artificial neural network altogether, which greatly improved the efficiency and accuracy of image retrieval.

Keywords-Artificial Intelligence; Image Semantics Retrieval; Image Segmentation; Image Annotation

1. Introduction

Along with the rapid development of multimedia technology and the Internet, images have been widely applied in all industries, making image resource retrieval increasingly significant. Image semantics retrieval has become a research focus at present. The technology conducts retrieval through the logic features and abstract properties of images, improving computers' ability in image retrieval closer to human beings' comprehensive ability. However, how to extract semantic features is still under exploration and research [1]. Therefore, we proposed the idea to solve the issue of semantic feature extraction in image retrieval by making use of artificial intelligence, which would in return strengthen computers' ability in image retrieval, making it closer to human beings' comprehensive ability.

2. Analysis

At present, there are mainly three types of existing image retrieval technologies, which are separately based on text, content, and semantics. The test-based image retrieval makes note of images manually, and then carries out keyword retrieval by making use of the text retrieval technology. However, this technology is not pragmatic at all. First of all, we'll have to make note of all images manually and completely, which might be easy for small image sets. Yet, along with the increase of image quantity, especially that there are large amount of images in the Internet, it would be a tremendous workload to make note of all images manually. Obvious, this retrieval method is never an ideal choice. Secondly, as images contain vast amounts of information, while different users may hold different opinions over the same image, so that the subjectivity and fuzziness in image annotation will inevitably affect the subsequent retrieval process. There is no unified standard for image annotation. Therefore, it will be difficult for the retrieval result to perfectly comply with users' requirements. Content-based image retrieval, i.e. CBIR, is part of image analysis, which is designed to search and figure out images that meet the requirements from the image database in accordance with the content information or the specified query standard based on the given target image. In the retrieval method, the query condition is actually an image or a description of the image content in itself. Correspondingly, it builds index by extracting the bottom features, and then makes comparison of the differences between these features and the query condition through calculation, so as to work out the level of similarity between the two images. This retrieval method relies less on manual intervention, and has been widely applied in many fields, leading to many distinguished achievements in automation applications. Although the method can be deemed to be a mainstream technology in image retrieval, it still has a long way to go before reaching the aims of the ideal high efficiency and accurate retrieval. Because of the deficiencies in the aforementioned retrieval methods, people are attempting to find an intelligent retrieval method of high level. As people's judgment over the similarity of image is not just based upon the visual features of image, which further consists people's comprehension of image contents. Such comprehension can never be acquired directly from the visual features of image, but from people's knowledge. This comprehension of image contents aforementioned is just the semantic feature of image. Only by integrating various information of image, especially the semantic information, can retrieval system's ability in image retrieval approach as closer as possible to human beings' comprehensive ability. For this reason, images shall be affiliated with all kinds of information, including semantics, so as to fully meet users' needs. Semantics-based image retrieval conducts retrieval according to the logic features and abstract properties of image, which would in return strengthen computers' ability in image retrieval, making it closer to human beings' comprehensive ability. At present, image semantics retrieval has become a research focus in the field.

3. Methods

The three research objects of artificial intelligence are fusing fuzzy logic, genetic algorithm and artificial neural network. Artificial neural network has the ability of study, association and fault-tolerance, and is adept in studying directly from data. Fuzzy system has strong reasoning ability, and is good at describing and utilizing knowledge in disciplinary areas. Genetic algorithm has several intelligent features, including self-organization, self-adoption and self-study. This method is expert in solving global most optimal problem. We fuse these three methods together to realize the extraction of semantic features in semantics-based image retrieval.

4. Implementation

4.1 Image Segmentation

Image segmentation separates relevant targets from the background, and then makes a classification of all pixels. The basis of classification can be the grayscale of pixel, the size, speed, color of target, as well as space, texture characteristics, etc. As for this, we can make use of the multi-characteristic parameter coding gene cluster for an optimized segmentation. In genetic algorithm, the length of chromosome is determined by the actual requirement of the problem to be solved. Decoding will also be carried out according to the meaning that every cluster indicates, so as to figure out the global optimal value through genetic algorithm [2].

a) Build an initial group constituted by character strings. In image segmentation, set the population as 20, the number of generation as 40. As the grayscale value of the segmentation electrical level varies between 0 ~ 255, so that we can encode all chromosomes with binary strings of 8-digit numbers of 0 and 1, which indicates a certain threshold value. The initial values of these thresholds are random, with both higher and lower fitness correspondingly. Following that, we will decode the chromosomes with genetic algorithm, as well as working out its value of fitness. The value of fitness will be adopted to make selection of chromosomes through the test of survival of the fittest. The chosen chromosomes will serve as the parent generation [3]. After a constant evolution, chromosomes will converge in the last generation with their fitness reaching the best, which is the best threshold.

b) Design of fitness function. Fitness function is equivalent to biological evolution environment in real world. The quality of environment will directly affect the result of biological evolution. Here, we employed the Otsu method as the fitness function in image segmentation:

$$f = w_0 (u_0 - v)^2 + w_1 (u_1 - v)^2$$

In the function, w_0 and w_1 separately stand for the probability when its grayscale is less than the threshold and larger than the threshold, while u_0 and u_1 respectively represent the average grayscale of the two intervals aforementioned, and v signifies the average grayscale of the entire image. The larger the value of is, the better the segmentation quality is.

c) On the basis of genetic probability, the following operations will be conducted to generate new groups. Here, there manipulators are included, such as duplication, simple intersection and bit variation. The manipulator parameter will be set as the intersection rate 0.65 and the bit variation rate as 0.008. Operand load, i.e. the population, will be set as 20. In a simple intersection, the under-mentioned method will be employed to re-group two parents so as to form two daughters, which enable favorable new features to combine immediately. Simple intersection is performed based on a certain probability. The larger the probability is, the bigger the possibility of intersection performance is.

4.2 Automatic Image Annotation

Generally, when retrieving an image, users will care more about the semantic features if the image. For instance, when retrieving an image with “tiger” in it, users would hope that all kinds of backgrounds and “tigers” of various patterns be retrieved out. However, seeing from the angle of feature, the features of images under different backgrounds are quite different as well [4]. Consequently, we adopted feedforward neural network for training and study. The training method is based on minimization of the pre-given evaluation function, which actually approaches the projection of the already built sample and its classification through a combination of several functions (i.e. output function in the hidden unit) pre-ascertained with form and quantity. Taking samples separated out with genetic algorithm in the last section as the initial samples, we then approach to its distribution in the space with the sample data its self, as well as to build the neutral network on this basis, and then “sketch out” approximately the geometry areas of all samples with combinations of neuron-covered areas. Moreover, when determining the classification of a new sample, we just need to figure out the membership degree of sample to these geometry areas, while the classification corresponds with areas of larger membership degree is just the answer.

We have introduced keywords to describe the semantics of an image. However, many keywords with similar semantics can be used to describe the same image, bringing about difficulties to the matching and management of keywords. For this reason, for the description of an image, we shall employ keywords of two levels. Firstly, classification information of an image will be utilized to indicate its semantics of the first level. For images with ambiguous semantics in the system, we can divide then into several categories, and then employ several keywords to indicate their first level semantics. As for the first level keywords, we can put to use some widely-accepted classified catalogues, so as to impose restrictions on the keyword volume in the level. In the second level of semantic description, users will then be able to employ various keywords for image description. By taking this strategy, we can to some extent overcome the diversity of keywords that caused by synonyms or other factors.

In order to obtain the semantic information of an image, corresponding keywords and weight of images shall be modified in accordance with users’ feedbacks. The process includes:

a) In system initialization, set corresponding keyword weight of every image as 1, indicating that all corresponding keywords of the image have the same significance.

b) After each query and feedback of users, gather the query keywords given by users, as well as the relevant and irrelevant feedback images.

c) For every query keyword submitted by users, we shall firstly search from the keyword database to ensure if there is corresponding keywords. If the answer is no, we then establish corresponding items, as well as corresponding links in the keyword database.

d) For all relevant feedback images, we shall firstly check if there are user-submitted query keywords in the keywords associated with every image. If so, increase the keyword weight with 1. Or else, add the query keyword into the image database and then endow the keyword with a weight of 1.

Append new images with semantics annotation strategy:

a) Store all new images to be annotated into a specified catalogue. After these new images have obtained semantic information through users’ feedbacks, we then move the image out of the catalogue, so that we can visually acquaint ourselves with the semantics annotation progress of new images in the image database.

b) For every new image in the catalogue, we will firstly calculate its visual similarity with every image in the image database, and then collect the keywords of the first n most similar images, as well as take the preceding keywords as the keywords of new images, with a weight of 0.5.

c) The keyword weights of new images shall be checked during every feedback. If the weight of a certain keyword. has exceeded the given threshold value, the keyword will then be taken as an ascertained image keyword. As for this, other keywords estimated in *b)* shall be deleted, and the image shall be move from the new image catalogue into other specified catalogue.

5. Conclusion

Fussy logic shows strong reasoning ability in dealing with ambiguous fussy information, which further reveals human beings' experience and general knowledge. Moreover, neural network has strong study ability. By integrating the two parties altogether, we can acquire strong structural knowledge expression ability with fussy logic, as well as the neural fussy network with strong studying ability from neural network. Owning to the intrinsic characteristics, genetic algorithm can adopt random and oriented retrieval principles to obtain the most optimal solution. Consequently, in the structure of neural fussy network, genetic algorithm can play an important conducive role. Therefore, we put forward that the fusion of the three can be applied to realize semantic feature extraction in semantics-based image retrieval. Furthermore, system built on this basis has the ability to expand dynamically the knowledge in semantics-based image retrieval, as well as the competence to acquire actively semantics features of images.

6. References

[1] WANG Hui Feng; SUN Zheng Xing; and WANG Jian, "Semantic Image Retrieval:Review And Research**,** Journal of Computer Research and Development,vol. 5, pp.2-12, 2002.

[2] ZHANG Lei, LIN Fu-Zong and ZHANG Bo, "A Forward Neural Network Based Relevance Feedback Algorithm Design in Image Retrieval," Chinese Journal of Computers, vol. 7, pp. 2-9, 2002.

[3] YANG Hui, QU Xiu Jie, "Survey of Image Segmentation Method**,"** Survey of Image Segmentation Method, vol. 3, pp. 22-24, 2005.

[4] SHI Yue-xiang, ZHU Dong-hui, CAI Zi-xing and B.Benhabib, "Extraction of Image Semantic Attributes and Its Application," Computer Engineering, vol. 19, pp. 183–185, 2007.