A Method of the Semantic on Image Auto-Annotation

Huo Lin[*], Liu Xianwei and Huang Baohua

School of Computer and Electronic Information, Guangxi University, Guangxi, Nanning, 530000, P.R. China

Abstract: Recently, due to the existence of semantic gap between image visual features and human concepts, the semantic of image auto-annotation has become an important topic. Firstly, by extract low-level visual features of the image, and the corresponding Hash method, mapping the feature into the corresponding Hash coding, eventually, transformed that into a group of binary string and store it, image auto-annotation by search is a popular method, we can use it to design and implement a method of image semantic auto-annotation. Finally, Through the test based on the Corel image set, and the results show that, this method is effective.

Keywords: Color correlograms, hash code, image auto-annotation, image retrieval.

1. INTRODUCTION

With the rapidly development of the Internet, as well as the arrival of the era of big data, the number of data exploded with the digital camera which requires a management systems. Big data on the structural features can be divided into structured and unstructured data. In the unstructured data, with respect to the text data, the information contained in the image is more complex. In many areas such as the news, media, medicine, and other entertainment, in order to effectively organize, query and browse such a large number of image resources, image retrieval technology came into being, and it has become a hot research topic in recent years.

However, compared to the text retrieval technology, which is mature and widely used, the large-scale image retrieval technology is in the exploratory stage. The crucial issue is the semantic gap that between image visual features and human concepts [1-3], and how to solve the problem of the "semantic gap", which is a technical problem that the content-based image retrieval system is urgent to solve. And because the image annotation (refers to use the semantic keywords to represent the semantic content of an image) can effectively compensate for the lack of semantics, as well as to build a "bridge" between high-level semantic features and the underlying visual features, and then, through the image annotation, we can change the problem of image retrieval technology into text retrieval that is relatively mature. However, since manually annotating images is a very tedious and expensive task, so image auto-annotation has become a hot research topic in recent years. So we will discuss this topic systematically and comprehensively.

In this paper, the second section will discuss the related works of the authors introduced. In section 3, we will introduce the corresponding model and structure of the method that we have provided. We have made some test based on the

method that we have introduced in the section 4. In the last section, we give some possible future works and the disadvantages of this paper on the image semantic auto-annotation.

2. RELATED WORKS

Currently, the topic about the Semantic on Image Auto-Annotation has been widely studied [1-6]. The method is used to annotate image in various way. Roughly speaking, recent approaches mainly work on three directions: Category-based automatic image annotation, automatic, probability model based automatic image annotation, and image retrieval based using data-driven automatic image annotation. The method of Category-based automatic image annotation, is trying to solve the problem from the perspective of classification. Which is assuming that each of semantic keyword as a category of tag. The representative method of it is support vector machine (SVM), a mixed classification model, Bayes Point Machine, as well as 2D-MHMM model etc [4-6], these methods are set through a lot of training to learn a set of multiple classifiers, and by use the classifier model to annotate image, and set the corresponding classification results to the image as its semantics. The method of Probability model based automatic image annotation, is trying to solve the issue by analyzing and calculating the relevance or the joint probability between the image visual features and semantics key words. The representative method in the literature [7-9], in which two kinds of probabilistic models that mostly representative and used is probabilistic latent semantic analysis referred PLSA [10] and latent Dirichlet allocation referred LDA [11]. With the advent of large scale image database, the method of research-based image annotation is emerging, this method is used data-driven based to establish the contact between the image and text [12-14].

In the actual task of image annotation, the semantic concept need to be annotated is very much, that is meaning the method of category-based automatic image annotation need to train a large number of classifiers, it's difficult to promote in large-scale, multi-class, real-time image annotation sys-
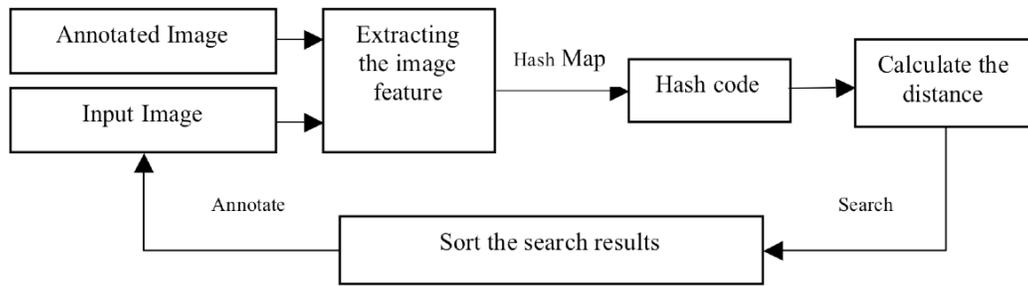
Fig. (1). The block diagram of designation.

tem. Although the method of Probability model based automatic image annotation have a good mathematical model to describe. But in the learning process of the model, there will be a complex parameter estimation and probability inference that makes the method difficult to apply to large-scale image annotation. However, the method of research based image annotation, there isn't need model training, as well as independent to model, thus better to avoid the difficult to promote since the disadvantages of restricted training samples and training time, has become an effective way of large-scale image annotation.

Sketch rebuild technique is the key point in making a 3D modeling from designer's 2D and 3D design sketch, its development is relativity mature. There are some product concept shape fast generating systems and CAD system that based on the sketch. Some big CAD/CAM software systems could provide corresponding module.

This paper will focus on the method of retrieval based image annotation, and the method of retrieval based image annotation is urgent to solve two major problems: first is how to define the distance that can measure its corresponding semantic similarity between the visual feature of two image, secondly, we need a large enough image library that is annotated. And here, we assume that we have enough data that is annotated, and the other way, we adopted the visual features of image are mapped to hash code, and by comparing the Hamming distance between hash code to achieve the judgment of similarity between semantic [16], thus achieving a method of the semantic on image auto-annotation.

3. THE CORRESPONDING MODEL AND STRUCTURE

The method that this paper proposed consists of three key modules, the visual feature of image extraction module, the similarity determination module, image auto-annotation module. The block diagram of designation will be shown in Fig. (1).

3.1. The Model of Image Semantic Auto-annotation

Generally, the image semantic auto-annotation can be described in the language of probability, which essentially is find w in the range of the key words to make the conditional probability $p(w|I_i)$ to be maximization. As shown in (1), where I_i denotes the input image, and w is all the candidate words in the annotated dictionary. Assuming that we have a large image database that is annotated, so that we can

make the (1) to expand the (2) by take advantage of the Bayesian formula, where I_q denotes a image that is similar to input image in a large-scale database, while $p(I_q|I_i)$ means the similarity between it and the input image, $p(w|I_q)$ means that the semantic correlation between the candidate vocabulary and image I_q. And further, assuming that there have a hidden layer between the image and the text(commonly referred to as the theme layer), we will use the appropriate distribution of probability to describe every image, and to describe these themes through the appropriate text, so we can further the (2) expands to (3) by using the total probability formula, where the theme of space denoted as t, $p(t|I_q)$ represents the matching between the theme of space and image I_q, and $p(w|t)$ indicates the accuracy of the topic space t described by the candidate vocabulary w. As the literature [7-9], achieving the image semantic auto-annotation by using the topic model, just as denoted by (3), it's a three-tier model structure of image semantic auto-annotation. While this paper based on the method of search-based image semantic auto-annotation, shown as (2), we have designed a two-tier structure of image semantic auto-annotation.

$$w^* = \arg\max_w p(w|I_i) \tag{1}$$

$$= \arg\max_w \sum_{I_q} p(w|I_q) * p(I_q|I_i) \tag{2}$$

$$= \arg\max_w \sum_{I_q} \sum_t p(w|t) * p(t|I_q) * p(I_q|I_i) \tag{3}$$

3.2. Image Feature Extraction and Mapping

The low-level visual features of image are mainly color, texture, shape, spatial relations etc, the visual feature now widely used is color features, mainly in color feature with a more clear definition, and its feature extraction method is relatively simple, the objects contained in the image and the scene can be preferably reflected by the color features. This paper uses 36 color correlograms as the original visual features [15], which is widely used for content-based image retrieval (CBIR).

In the above, since extracted the color correlation matrix, if the color characteristics of the image was directly to be stored, then we matching a input image, will need to search a large number of image, it takes a lot of time to compared,

assuming that we firstly further the color feature into the hash code and then to be stored, that's meaning each image is t to be mapped to a set of hash code, this would save a lot of space and search time, thereby improving efficiency.

After obtaining the above-mentioned color correlation matrix, due to the high dimensions of the matrix, so we need further dimension reduction of its operation, one case in order to retain the original information as possible, to get a more compact representation, second is to remove some small amount of noise, and does not affect the correlation value. Assuming that the correlation matrix obtained from the above denoted F_i. the dimension reduction can be accomplished by using a mapping matrix A, $G_i = AF_i$ to achieve, where the matrix A can be obtained by make the main component analysis on a large number of image data sets (PCA), and for all of the images, the matrix A is the same. Finally, with PCA operation, the 36-bin color autocorrelation matrix can be mapped to 32-dimension hash codes.

The hash code generation is essentially a vector quantization (VO) process, the quantization strategy is that if a feature component is larger than the mean of this vector, it is quantized to 1, otherwise to 0, as shown in (4),

$$H_{i,k} = \begin{cases} 1 & , & G_{i,k} > mean_k \\ 0 & , & G_{i,k} \leq mean_k \end{cases} \tag{4}$$

where $mean_k$ is the mean value of dimension k, $G_{i,k}$ denotes i-th element of the dimension k, in this way, the k-dimension feature vector is transformed into a k-bit binary string.

3.3. The Similarity Measure of Image Features and Auto-annotation

The similarity measure of image features normally after the image feature has been quantified, and the quantification of image features that directly operate mainly to multidimensional matrix after extract the image feature, therefore, the method of similarity measure for image feature is also more currently, such as the absolute value of the distance, Chebyshev distance, Euclidean distance, Mahalanobis distance, Gram distance etc. while in this paper, in view of taking into account the characteristics of the image was extracted, and further hash coding operation was carried out, thus the similarity measure used in this paper is the Hamming distance. It calculated as follows,

$$\sum_{k=1}^{L} H_{i,k} \oplus H_{j,k} \leq T \tag{5}$$

As shown in (5), compared the similarity of the two image's low-level feature, and convert it to compare between binary strings, where the parameter T can be designated by a large number of experiments.

Thus, the method of image semantic auto-annotation can be as follows: first, extract the low-level feature of the input image and annotated image, then reduce the image feature dimension and take the hash coding operation, calculated the

Hamming distance of the input image and the image sets, finally, sorting the Hamming distance by small to large order, assigned the semantic keyword of the image that has the minimum Hamming distance compared to input image to input image.

4. EXPERIMENT AND RESULTS

This experiment used the Corel5K image library as a training set and test set. Which contains 5000 images, covering natural scenery, people, animals, plants etc, in which 4500 images as the annotated image library, 500 images as input images, there are 374 words tagging in the words library, and each image is labeled 1-5 words as mark words. Semantic auto-annotation performance can be used recall and precision to evaluate, the definition of recall and precision as the following formula.

$$Recall = \frac{the\ results\ related\ to\ properly\ label}{all\ relevant\ results} \tag{6}$$

$$Precision = \frac{he\ results\ related\ to\ properly\ label}{all\ annotation\ results} \tag{7}$$

This experiment run in memory 2GB, Windows7 operating system, and use the software MatlabR2011b version, complete testing research, tagging performance test results shown in Table **1**.

Table 1.   Properties Performance on this paper.

Algorithm	Evaluation Criteria/%	
	Precision	Recall
This paper	65.93	94.54
Literature [17]	16.14	17.51

5. CONCLUSIONS AND FUTURE WORK

This paper designs and implements the method of image semantic auto-annotation, based on the color correlograms as the low-level feature, and then be in a specific way Hash coding, and ultimately form the image into a set of binary strings to be stored, by compared the distance between the binary string, we realized a method of retrieval based image semantic auto-annotation. This method is carried out in Corel image sets that contains 5000 images, and finally the experimental results show that the method is effective.

Since the method is using the color correlograms that now widely used for content-based image retrieval as the low-level feature, the semantic be annotated to image can only be limited the scene semantic, and for the high-level behavioral semantic can't be better identified,. Thus the future research word will be focus on to achieve a higher level semantic auto-annotation by combined with a variety of low-level features.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Z. Rong, "Research on semantic-based web image classification", *Ph. D. thesis, University of Zhejiang*, 2011.

[2]   R. Zhao ,W.I. Grosky, "Narrowing the semantic gap-improved text-based web document retrieval using visual features", *IEEE Transactions on Multimedia*, vol. 4, no. 2, pp. 189-200, 2002.

[3]   P. Enser, C. Sandom, "Towards a comprehensive survey of the semantic gap in visual image retrieval", *Lecture Notes in Computer Science*, vol. 2728, pp. 163-168, 2003.

[4]   C. Cusano, G. Ciocca, and R. Schettini, "Image annotation using SVM", In: *Proceedings of Internet Image IV, SPIE,* vol. 5304, pp. 330-338, Dec. 2003.

[5]   G. Carneiro and N. Vasconcelos, "A database centric view of semantic image annotation and retrieval", In: *SIGIR '05 Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* 2005, pp. 559-566.

[6]   G. Carneiro and N. Vasconelos, "Formulating semantic image annotation as a supervised learning problem", In: *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 163-168, 2005.

[7]   F. Tian, X. Shen, and X. Liu," Large scale image semantic relevance automatic annotation", *Journal of Computer-Aided Design & Computer Graphics*, vol. 25, no. 2, pp. 160-174, 2013.

[8]   J. Zhang, W. Hu, Z. Chen, and Y. Yuan, "Multi-model fused framework for image annotation", *Journal of Computer-Aided Design & Computer Graphics*, vol. 26, no. 3, pp. 472-478,2014.

[9]   Z. Li, Z. Shi, Z. Li, Z. Shi, "Automatic image annotation by fusing semantic topics", *Journal of Software*, vol. 22, pp. 801-812, 2011.

[10]   T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning*, vol. 42, pp. 177-196, 2001.

[11]   D.M. Blei, A.Y. Ng, and M.L. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.

[12]   X.J. Wang, L. Zhang, F. Jing, and W.Y. Ma, "Annosearch: Image auto-annotation by search", In: *IEEE Conference on Computer Vision and Pattern Recognition* 2006 pp. 1483-1490.

[13]   X.J. Wang, L. Zhang, X. Li, and W.Y. Ma, "Annotating images by mining image search results", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1919 - 1932 2008.

[14]   X.J. Wang, L. Zhang, X. Li, and W.Y. Ma, "Image annotation by large-scale content-based image retrieval", In: *MULTIMEDIA '06 Proceedings of the 14th annual ACM international conference on Multimedia,* pp. 607-610, 2006.

[15]   J. Huang, S. R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih, "Image indexing using color correlograms", In: *1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1997, pp. 762 - 768.

[16]   X.J. Wang, L. Zhang, X. Li, and W.Y. Ma, "Large-scale duplicate detection for web image search", In: *2006 IEEE International Conference on Multimedia and Expo*, 2006, pp. 353 - 356.

[17]   X. Zhang, M. Zhang, and J. Li, "Image automation annotation based on Matlab," *Modern Electronics Technique*, vol. 37, pp. 73-75, 2014.