

A Novel Image Semantic Understanding and Feature Extraction Algorithm

Xinxin Xie^{1, a} and Wenzhun Huang^{1, b*}

¹School of Information Engineering, Xijing University, Xi'an 710123, China

^a346148500@qq.com, ^bhuangwenzhun@xijing.edu.cn

*The corresponding author

Keywords: Image semantic; Algorithm; Feature extraction; Semantic understanding

Abstract. In this paper, we propose a novel image semantic understanding and the feature extraction algorithm. Image stitching algorithm based on feature is extracted from the image of basic certain characteristics as matching primitives, usually in the form of general dot, line, area, or some special structures, etc., rather than directly using the image gray level itself, and then to these features as a model for registration. Thus, because the image matching algorithm based on feature extraction and image just some of the aspects of characteristics, namely extraction pixel is less, and makes the image registration time is greatly reduced. Under this theoretical support, we proposed the image semantic understanding and the feature extraction algorithm with the experimental verification that proves the effectiveness of the method that is meaningful.

Introduction

In the theory of computer vision, image segmentation, feature extraction and basic target recognition constituted by the lower to the top three major tasks. Target recognition and feature extraction in the image segmentation as the foundation, the image segmentation result is good or bad will directly affect the subsequent feature extraction and target recognition [1-3].

In recent years, with the new challenge scene classification is put forward, the variety of the scene classification techniques emerged. In order to more clearly reflect the core development of the scene classification techniques that shows as much as possible the existing classification technology of the whole scene. The technical evolution process in the scene classification can be roughly divided into the three stages: from the original method, based on the underlying feature gradually transition to the method based on high-level features, the development to the present based on the characteristics of the learning. At each stage of different period also appeared a variety of different scene classification method, which based on high-level feature classification method and can be divided into the method based on scenario model and the method based on semantic scene.

To be sure, this is just a scene classification technology evolution trend, does not represent a row behind the method of one stage must than the previous phase of the method is effective. For example, now new scene of underlying characteristics still scene classification is an important breakthrough, and based on the deep learning method is not necessarily in some simple are very effective on small data sets. Automatic image annotation is a very challenging work in the area of the image retrieval research, is the key to realize image semantic retrieval. If can realize automatic image annotation, the image retrieval problem can be converted into the mature text retrieval problem, the automatic image annotation in computer vision, machine learning, information retrieval and other aspects of content which has a strong research value [4].

According to the literature review, the state-of-the-art image semantic understanding algorithms can be summarized as the listed aspects. (1) The automatic image annotation algorithm based on the probability correlation model. Intuitively, if two images with high visual similarity, they mark the higher the probability of the similar keywords sequence. The characteristics of this approach is that without learning mechanism for each semantic keyword set up a corresponding said low-level visual features. (2) Based on the learning algorithm of basic automatic image annotation. Based on mapping algorithm is a kind of a semi-supervised learning algorithm, a known class mark training data and the

unknown class mark test data will be involved in the learning process of the algorithm. And compared to the traditional supervised learning and unsupervised learning, a semi-supervised learning can use more information in the learning phase, such as the distribution of data characteristics, it is suitable for according to the total amount is larger, a relatively small quantity of the labeled training data. (3) Based on the classification algorithm of the automatic image annotation. Existing labeling algorithm based on classification are simply from the angle of the sample learning more or learn more marks to describe and solve the problem of annotation. Although these methods have different characteristics in the specific expression, but their core idea is consistent, namely using the known annotation data to establish a model to describe the potential connection between text and image characteristics of the vocabulary or mapping relationship, and predict the unknown image annotation.



Figure 1. The Image Semantic Understanding Database

In this article, we propose a novel image semantic understanding and feature extraction algorithm. This paper presents a new image ontology annotations framework. We will image semantic concept and high-level abstractions can be defined as attributes, quadratic mark method was adopted to realize automatic tagging for image semantics. In the mark for the first time, then give the building area based on Bayesian network object classes and properties of the ontology concept, the concept of relationship and combining with the image of the ontology concepts, logical semantic relationship between for the secondary labeling method. In the later sections, we will discuss in detail.

The Proposed Algorithm

The Principles of the Image Semantic. Visual language model is a model of analysis of the image block space information. An image can be divided into many areas, each area may, in accordance with its visual characteristics expressed in a visual words. Not only is there the spatial dependence between these visual words, there are a total of frequency by using these features, between visual words can easily dig image semantics [5-6].

Image semantic mining research focus in the field of the image classification is based on image semantics, in which the image semantic mining of human cognitive behavior, its core is that from an existing in the training data to high-level semantic concepts and the relationship between low-level visual features of image and the automatic modeling, using the learned model for automatic mining unclassified image semantics. By estimating the conditional probability distribution of visual words can obtain image block interdependence between semantic relations as the follows.

$$p(w_i | w_{i-1}, C_t) = \frac{F(w_{i-1}, w_i | C_t)}{\sum_{\Theta} F(w_{i-1} | C_t)} \quad (1)$$

In image semantic mining, said image semantic dimension of feature space is generally high, at the same time there are a lot of synonymous visual words and visual word polysemy, the existence of the redundant information, not only lead to the distribution of the model based on the visual language of basic visual words calculation result is invalid, and cannot be used for image semantic mining well for automatic classification of image semantics. As the satisfactory tool for analyzing the corresponding issues, the SVM can be introduced for optimization.

The SVM classification performance is affected by many of factors, of which the following two factors is the key. (1) Form of kernel function. Different kernel functions lead to the generalization performance of SVM is different, so how to choose according to the specific data of kernel function is directly related to the performance of the SVM classification. (2) The error penalty parameter C and kernel functions parameters. In determining the feature subspace of C value of small said the penalties for experience error is small, the complexity of the learning machine small and empirical risk value; C in infinity, all constraint conditions must be met, this means that training sample must be classified.

Technology of application of the kernel function, the nonlinear problem of input space, through function mapped to high-dimensional feature space, linear discriminant function in high-dimensional space construction, commonly used kernel function has the following several categories.

$$k(x, x_i) = (x \times x_i) \leftrightarrow \text{linear} \quad (2)$$

$$k(x, x_i) = \exp(-\lambda |x - x_i|^2) \leftrightarrow \text{RBF} \quad (3)$$

The Image Feature Extraction. Image matching technology is to realize the image fusion, image correction, image Mosaic, and one of the key steps, target recognition and tracking has been widely used in the field of image recognition and image reconstruction, etc.

Image matching for different source image matching, is difficult in the infrared image and visible light image are the most common and often used in military field, and visible light and infrared image imaging mechanism and different shooting conditions, there are large difference between the edge of the image contour, even very fuzzy, noise is bigger, therefore, the research of the high precision, fast correct matching rate is high, anti-jamming strong matching algorithm has the very vital significance.

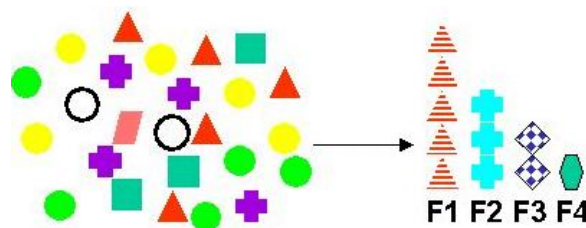


Figure 2. The Image Feature Extraction Procedure Demonstration

In order to overcome the disadvantage of image matching method based on gray level, people put forward the method based on feature, based on the characteristics of the method first extracted from to match images contain important features of image feature points, with the similarity measure between image feature point matching that is commonly used in basic image matching features such as edges, outline, linear and angular point. Matching based on feature for image distortion, such as block has certain robustness, but its matching performance depends largely on the quality of feature extraction, and this kind of method of feature extraction is easily affected by noise [7].

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right) \quad (4)$$

Scale invariant feature transform is a method of feature detection based on the feature description. The SIFT algorithm is zoom, rotate and affine invariability, can resist some illumination changes and perspective transformation. At the same time, this characteristic also has the high distinguish ability, is advantageous to the subsequent matching. The algorithm firstly using the feature point detection method to extract the two corners of the image, and secondly by the Taylor series to eliminate pseudo angular point, in turn, determine the feature points, to determine the core feature points on each curve calculating the gradient vector and direction angle correction, then the feature points neighborhood part adopts nearby projection principle, and the gradient direction histogram to determine the main direction of the feature points, finally in polar constraint bidirectional matching method was designed on the basis of feature point matching as the formula 5.

$$x = \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad (5)$$

At this time make the other side of the pixel candidate feature points, and then repeat the above calculation, get the new subpixel location precision. Finally, by using the subpixel accuracy position replace all scale candidate feature points location [8].

The Image Semantic Understanding. Parallel regional technology is image processing and analysis theory of the technique of image analysis. In the study of image and application, people often only interested in some of the images. These parts are often referred to as the target or prospects, the rest of the known as the background as they are usually in a specific area, the unique quality of image. In order to identify and analyze the target, need to separate out the relevant area, on the basis of this is possible to increase the use of the basic target, such as feature extraction and measurement, image segmentation is refers to divide the image into the distinctive regional and extract the interested target technology and process, this feature can be gray, color, texture, etc., the target can correspond to a single area that also can correspond to multiple regions.

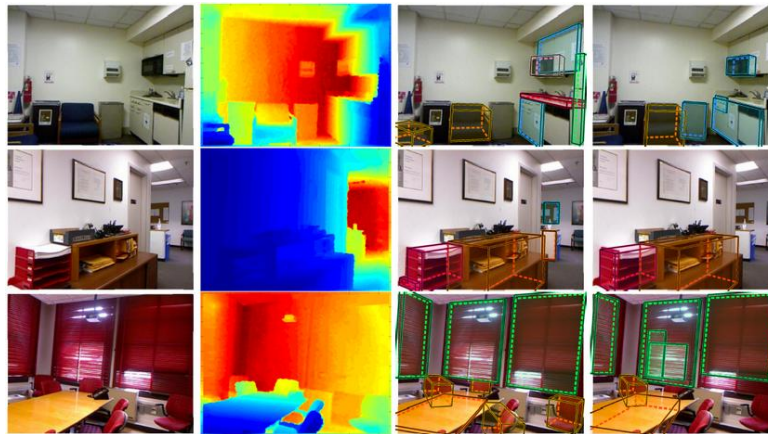


Figure 3. The Verification of the Image Scene Understanding

The complex scene of the video monitoring system, the tracking technology is not perfect enough, such as multiple target tracking of the occlusion may make track interrupt each other, and complex behavior variety, path length difference is big, therefore, is not suitable for the complex behavior of trajectory modeling directly. The complex behavior by landmark tracks down into simple behavior we can use the same simple behavior training as the follows.

$$O_{flow} = (O_v^1, O_v^2, \dots, O_v^T) \quad (6)$$

GIST features low computational complexity and easy to use, good at depicting the scene of the overall layout, but in the background of complex or the dense poor performance in the scene. SIFT features to translation, scaling, scale transform, brightness change, perspective transformation has certain stability, but not very well describes some of the overall layout of core complex scene. HOG features can be described on the edge of the shape and the target contour, but in the shape of the scene structure under the condition of that unstable failure. Characteristic to have with local features by describing the scene geometry information of scene, said the global but the characteristics of scaling and rotation invariance, and ignore the color information of the scene, so the complex scenario in some poor performances.

Images of the underlying features in simple scene classification task is a kind of effective, says but as a result of this representation are usually with a few semantic information, making them under the condition of the scene is more complex expression is very limited, so how to understand high-level semantic became a scene of scene classification new breakthrough. In view of scene and changeable situation, it removes the limitation of the whole space structure, in depicting the image area into the local spatial structure characteristics, the second stage is the classifier and the categorizing scenario testing completely separated, the image area for classifier to further optimize the weight parameters make the classification more effective. But it also has its own limitations, so that the trained model to different data sets is versatile of course this is also the most classification method has a defect.

Summary and Conclusion

In this paper, we propose a novel image semantic understanding and the feature extraction algorithm. When local content under the conditions of different imaging features of obvious differences that will lead to feature descriptor is difficult to match, for this, put forward a kind of image feature descriptor based on stable area. Maximum stable extremal region detection algorithm to extract image stability region, combined with the location and scale of image feature points, according to the characters of the stable area calculation principal direction, in stable region adapted to the feature points within the larger neighborhood scale, generated based on the characteristics of the logarithmic polar coordinates description operator. Under this background, we then apply the corresponding issues on the image semantic understanding tasks that assist to achieve the better performance.

References

- [1] X. M. Liu, H. X. Yao, R. R. Ji, P. F. Xu and X. S. Sun: Multimedia tools and applications, Vol. 64 (2013) No.1, p.53.
- [2] S. Gupta, P. ArbeláezArbelaez, R. Girshick and J. Malik: International Journal of Computer Vision, Vol.112 (2015) No.2, p.133.
- [3] S. Aditya, Y. Aloimonos, C. Baral, C. Fermuller and Y. Z. Yang: *2015 AAAI Spring Symposium Series* (Stanford University, California, March 23, 2015).
- [4] X. F. Wang and J. H. Li: *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering* (Hangzhou, P. R. China, March 22-23, 2013). p.1669.
- [5] T. D. Kelley and S. McGhee: *Proc. SPIE 8756, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2013* (Baltimore, Maryland, USA, April 29, 2013).
- [6] P. Molino and L. M. Aiello: SMIR@ SIGIR 2014 (Gold Coast, QLD, Australia). p.38.
- [7] A. Amnal, A. Dewabharata, S. Y. Chou and M. Erawan: *Proceedings of the Institute of Industrial Engineers Asian Conference 2013* (Springer Singapore, July 12, 2013). p.1081.
- [8] J. L. Martin: *To boldly proceed: Papers from the 39th International Systemic Functional Congress* (Edited by John S. Knox, Sydney, Australia, 2012). p.183.