ATLANTIS
PRESS

Research on the Image Semantic Understanding Pattern based on the Sparse Coding and Wavelet Theory

Wenzhun Huang[1, a] and Baohui Zhao[1, b*]

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

[a]huangwenzhun@xijing.edu.cn, [b]hui546837282@qq.com

*The corresponding author

Abstract. In this paper, we conduct research on the image semantic understanding pattern based on the sparse coding and wavelet theory. From the perspective of image understanding, according to the different weights of spatial information to give false points pixels, make the false points of pixels in different positions have different importance. Was introduced to the weighted formula of relative distance, foreground and background search distance, in order to obtain evaluation algorithm for image segmentation is scaling invariance conclusion in order to overcome large ratio of segmentation quality evaluation distortion phenomenon, put forward a kind of distortion of punishment, improve the effectiveness of the evaluation algorithm and global. Through applying the algorithm into primary application scenarios, we could test the effectiveness and robustness.

Introduction

Real image is varied and every image can correspond to one or more tags. Existing methods to image labeled as to the label of the mapping process, they are all the tags with the complete image, the latent semantic model, for example, cross-media correlation model, etc. In order to better understand and analyze the image, only the content of the image level is not enough that need more information of the image region level. Image tag can be label with a specific point that can also be a high-level semantic label using this algorithm can identify the specific object or the area that also can mark of high-level semantic information label before using the proposed algorithm for the basic image annotation to treat annotation image preprocessing. Pretreatment is included in the image and to join with artificial tag labeled with network image, the image segmentation, feature extraction. Then use of image labeling algorithm based on regional context awareness annotation image sets which can get each the image corresponding to the set of tags [1-3].

According to the literature review, the procedures of the image understanding could be summarized as the follows. (1) Identify area category. Ideally, the image segmentation can accurately extract all areas in an image, and then classifying regions tags, each area can get the corresponding labels. But in actual cases due to the limitation of the level of image segmentation, and can't get the complete area. (2) Clustering tags unknown area. Real image area are diverse, only rely on a limited number of basic classifier classification is difficult to meet the demand of annotation. For those who do not belong to the classifier recognition within the scope of the area, with a network of the artificial label image is introduced in this paper, to help mark the label. After get all the unknown area, through clustering can get a certain number of areas, in the collection contains many labeled network image region, and then use the label propagation way to annotate unknown area. (3) Tag correction. Conditional random field model is an undirected graph model, which point in this picture said, connected to the two side said the relationship between the two labels. Use the conditional random field model combined with the prior knowledge of the labels and tags can be fully the relationship between knowledge annotations.

Sparse method is the basis of de-noising: assuming that the image is sparse, natural signal can be represented by the dictionary atomic linear combination, and noise is not sparse. Therefore, the image sparse decomposition, under the over complete dictionary of sparse component is regarded as useful image signal, and the residual noise is considered. Remove the residual, and by the sparse coefficients

obtained after decomposition and dictionary of image reconstruction that can achieve de-noising and will assist for the further semantic understanding pattern analysis. Because it can effectively describe the structure of the scenario prior, local and global prior express scene and can effectively under the same energy optimization framework integration scenarios are means of understanding, high-order energy models applied in the understanding of the scene brought enormous advantage and potential but the high-order energy model also many inherent deficiencies [4].
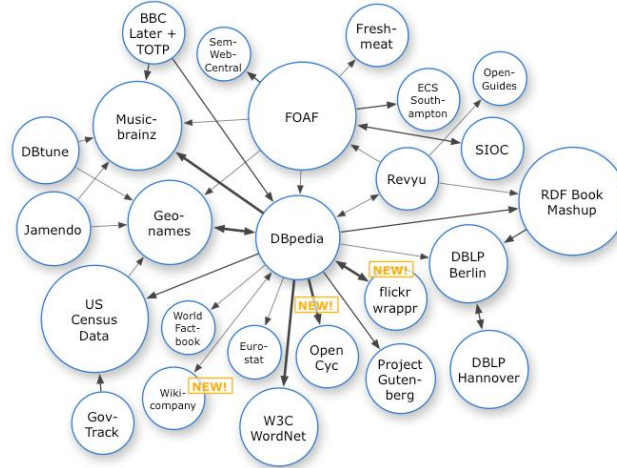

Figure 1. The Concepts of the Semantic Understanding

In this article, we conduct research on image semantic understanding pattern based on the sparse coding and wavelet theory that will be discussed in detail later.

The Proposed Methodology

The Sparse Coding. As a main representative of the compressed sensing technology, a sparse coding is to use some training can represent the low-level features of a linear combination of the base vector to represent the characteristics of the input image. Traditional dictionary to represent image features and classification tree method, is essentially every stay quantitative characteristics of the image, select from its nearest neighbor said one word as its characteristics, namely in a fixed term, use only contains a nonzero elements of the vector to represent a feature. This method feature of visual word near the border between quantization errors is bigger, and the quantization error is inevitable. Using sparse coding method can minimize the quantization error and it can stay by solving the convex optimization problem to calculate the quantitative characteristics of sparse representation of base vector, using the multiple linear combinations to represent a feature vector as the formula one.

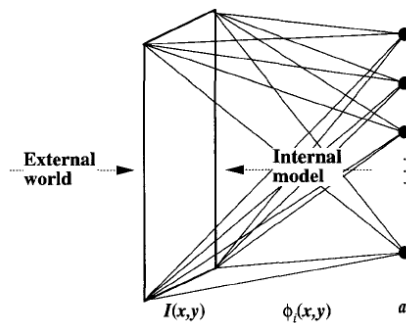$$x = \sum_{i=1}^{N} \theta_i \tau_i \tag{1}$$

Figure 2. The Signal Sensing and Obtaining Scenario

Error sum of the squares cannot well evaluate the image visual quality, because it does not give full consideration to the human eye visual features, the image that he eventually to the human eye's signal is very unreasonable. Natural image signal is highly structured, they have strong correlation between pixels, especially when these pixels in the space near the correlation carries the important structural information

of visual objects and the sum of the squares of the error only considering the different between image pixels, not considering the strong correlation between image pixels and the structure of the importance of visual object, for this, we revise the issues as the follows [5].

$$\delta_{x,y} = \overline{xy} = \frac{1}{N} \sum\nolimits_{i=1}^{N} xy_i \tag{2}$$

The change of node numbering corresponds to the ranks of matrix A of adjustment. Obviously, A and B are decomposed factor table after the central element number is different. Node optimal number algorithm is to seek an optimal matrix B, so that when the matrix decomposition factor. The decline of the non-zero will significantly reduce the number of matrix operations and necessary storage space, so as to improve the efficiency of the operation.
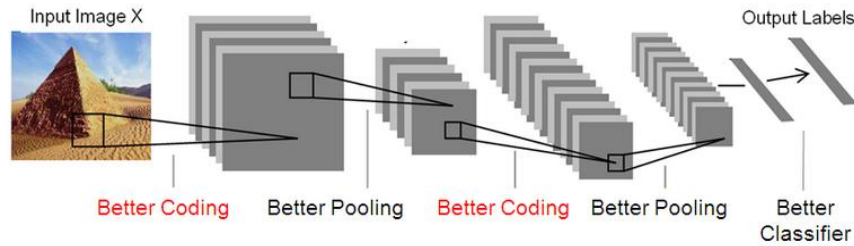

Figure 3. The Sparse Coding Procedures

The difference between sparse signal groups can use separate coding and distributed compressed sensing algorithm of joint decoding for decoding signal group as the formula 3.

$$T\left(optimized\right) = \min \frac{1}{2} w^t w + C_1 \chi + \sigma d \tag{3}$$

Sparse coding is a kind of feature representation and its goal is by solving a small amount of the unknown can represent the low-level feature information base vector coefficient as these coefficients are used to convert the base vector to represent the input characteristics of the linear combination.

The Wavelet Theory. Wavelet analysis is a kind of signal and information processing tools, in the time domain and frequency domain at the same time has good localization property, overcome the function of Fourier analysis in a single variable signal defects. It can be a signal is decomposed into independent part of the signal in time and the space, without loss of the information contained in the original signal, and can find orthogonal basis without redundant signal decomposition.

Wavelet transforms as a new mathematical tool, because of its good time-frequency localization characteristics and computationally simple, has been widely used in image processing. Image wavelet multi-scale representation of a pyramid structure, this for fast image matching from coarse to fine is very favorable. At the same time, the process of image wavelet multi-scale decomposition, and can be regarded as a kind of the automatic multi-scale edge extraction process, because in the high frequency part of image wavelet decomposition, energy mainly concentrated in the image gray scale changes, and these places are often the corresponding edge points. In addition, the wavelet multi-scale image representation also has direction that makes use of this characteristic can realize the multi-scale edge matching with directional selectivity. For instance, the wavelet transform is expressed as follows.

$$\begin{cases} A_j f = \left\langle f\left(x,\mathrm{y}\right), \phi_{j,n,m}\left(x,y\right) \right\rangle, \left(n,m\right) \in L^2 \\ D_j^1 f = \left\langle f\left(x,\mathrm{y}\right), \varphi^1{}_{j,n,m}\left(x,y\right) \right\rangle, \left(n,m\right) \in L^2 \\ D_j^2 f = \left\langle f\left(x,\mathrm{y}\right), \varphi^2{}_{j,n,m}\left(x,y\right) \right\rangle, \left(n,m\right) \in L^2 \\ D_j^3 f = \left\langle f\left(x,\mathrm{y}\right), \varphi^3{}_{j,n,m}\left(x,y\right) \right\rangle, \left(n,m\right) \in L^2 \end{cases} \tag{4}$$

Combination of wavelet transform and curvelet transform image interpolation algorithm, wavelet transform sparse representation of image texture composition features and curvelet transform cartoon ingredients sparse representation of image features, image interpolation problem can be converted to sparse constraint, the image reconstruction using iterative shrinkage projection method to solve the

problem, to achieve the image resolution enhancement. Then the known pixels with the pixel weight function is multiplied by the corresponding results are cumulative, the greater the weight of the pixels in the influence of the accumulation of pixels, the greater the final result of the accumulation weight balance, guarantee repair pixels to get the most close to the actual pixel values of the image. Impact factor and the pixel by pixel direction distance constraints of the influence factor that can accurate the reconstruction of image pixel values as much as possible [5-6].
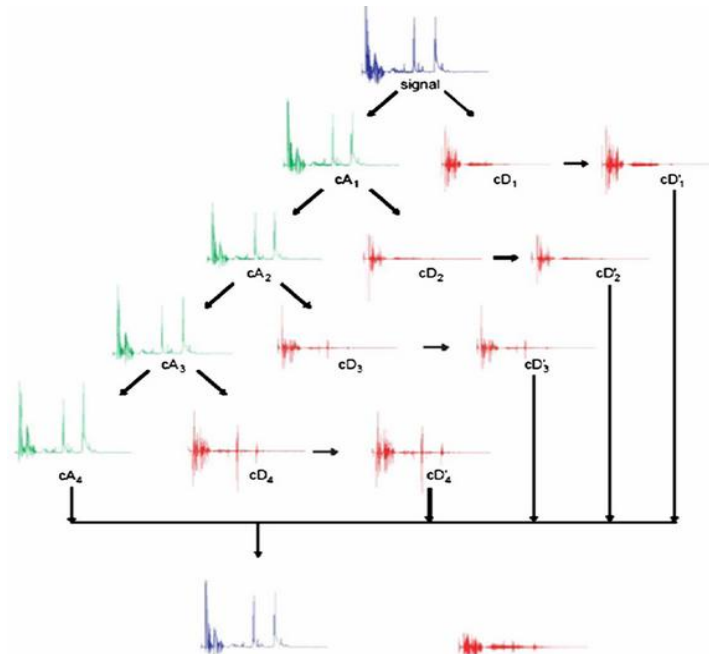

Figure 4. The General Pattern of the Wavelet Theory

The Image Semantic Understanding. Processing and analyzing relative mature algorithms of the digital image can provide us some powerful tools to understand some specific types of images, but to give a comprehensive image understanding method can work well on the all images is difficult, the main reason is the way for every image can understand a lot, and we know nothing to the things that can appear in the image. For the computer system, only by establishing corresponding to the expected object model to understand the contents of the image. Directly obtained image understanding is a target of a process, a single operating control as the focus, rough and detailed analysis must rely on the image correlation information, namely, relation between the current contexts as follows [7].

Segmentation knowledge. Based on the knowledge of image segmentation is to make full use of already accurately described in the system of the knowledge, giving guidance to the image segmentation, and the correct segmentation result is obtained. As a result, and problems related to the level of knowledge representation, determines the depth of image description, thereby to determine the degree of the image to understand itself.

Explain the knowledge. A scene understanding is essentially using corresponding relationship between the region, and interpretation model to quickly match to complete explain the process of target is obtained.

Regional knowledge. By using the relationship between the regional knowledge for image understanding can use an example to illustrate, if there is a chair in the image, then, under the condition, through the segmentation can be a face, four legs and a back of a chair, and then you can use it to the relationship between reasoning for identification.

Therefore, after getting classes in all areas of the probability, in this article, through the calculation area corresponding to the category of core uncertain degree to judge whether the area identified as a particular category that could be expressed as the follows.

$$E(s) = -\sum_{i=1}^{N} P(c_i \mid s) \times \log_2 P(c_i \mid s) \qquad (5)$$

Said in the diagram, the node area, edge said the distance between the areas is known and unknown area, it shows the unknown area and adjacent area. The relationship between the topology based on this structure that the algorithm can according to the two unknown area on the object level and the relationship between the surrounding area have much similarity to determine whether can be assigned to the same category. Therefore, we could obtain the separation of function as follows [8].

$$g(s) = \left[H_0(s), H_1(s), \cdots H_R(s) \right] \tag{6}$$

After getting the unknown context descriptor, combined with the feature of region of the visual information on the collection of clustering can get similar area. After getting cluster subset algorithm by using the advance with image centralized network image with user annotations to help mark the unknown area. Due to network image is processed with stay mark image, after clustering area in the collection of a certain area containing labeled images. Through the statistical method to determine the set of tags and then put these labels assigned to mark the unknown area.

Conclusion

In this paper, we conduct research on the image semantic understanding pattern based on the sparse coding and wavelet theory. Image an exponential growth with the development of Internet, network, image understanding technology becomes increasingly important. Image annotation technology as its key technology is widely attention and research. The existing image annotation techniques are mostly in the image tag model level training that ignores the relationship between the image area and the core relationship between the tags. In order to solve this problem, this paper proposed a new algorithm, combined with the location of primary relationship between regional and the symbiotic relationship between assistive labeled images. To be specific, algorithm firstly uses the wavelet analysis for some can determine area gives semantic tagging, and then take basic advantage of regional position help clustering with unknown area. In the later research, we will conduct experimental analysis for better analyzing the corresponding issues for further processing and optimization.

References

[1] H. Hajishirzi, M. Rastegari, A. Farhadi and J. K. Hodgins: Conference on Uncertainty in Artificial Intelligence (Catalina Island, United States, August 15-17,2012), p. 326.

[2] Jabeen: Academic Research International, Vol. 4 (2013) No. 6, p. 139.

[3] S. Aditya, Y. Z. Yang, C. Baral, C. Fermuller and Y. Aloimonos: Association for the Advancement of Artificial Intelligence Symposium (Palo Alto, California, USA, March 23-25, 2015), p. 9.

[4] J. Davey, S. A. Rueschemeyer, A. Costigan, N. Murphy, K. Krieger-Redwood and E. Jefferies: Brain and language, Vol. 142 (2015), p. 24.

[5] R. Girshick, J. Donahue, T. Darrell and J. Malik: IEEE conference on computer vision and pattern recognition (Columbus, Ohio, US, June 23-28, 2014), p. 580.

[6] B. Krausz and C. Bauckhage: Computer Vision and Image Understanding, Vol. 116 (2012) No. 3, p. 307.

[7] N. Goel and P. Sehgal: Intelligent Computing, Networking, and Informatics (Springer India, India 2014), p. 327.

[8] X. F. Wang and J. H. Li: International Conference on Computer Science and Electronics Engineering (Hangzhou, China, March 22-23, 2013), Vol. 3, p. 1669.