Smoothness Guide the Extraction of Multiple Resolution Patches from Image

XI ZHANG, SIJING DING, JIE WANG, YANGRUI RUAN, HONGJUN LI*
College of Science
Beijing Forestry University
Beijing 100083
CHINA
lihongjun69@bjfu.edu.cn

Abstract: - Extraction of multiple resolution patches from image is helpful to image semantic segmentation and scene understanding. Based on the analysis of the algorithm for structure extraction which realized via relative total variation and its parameter value, a semi-automatic algorithm for the extraction of multiple resolution patches is presented in this paper. We put forward a piecewise function of iteration step length and give an optimal stop criterion for the iterative algorithm. Our experiment results show that our method is of general applicability and feasibility.

Key-Words: - Multiple resolution patches, Extraction of patches, Image segmentation, Optimal partition, Annotation, Semantic segmentation, Smoothing.

1 Introduction

The emergence of mass image data in recent years brings an urgent requirement of the fast intelligent understanding of scenes. Image segmentation contributes to image processing, analysis and understanding. Extraction of multiple resolution patches from image is an expression of multilayered image segmentation, which can help to understand images from the view of hierarchical structure. For instance, De Smet V et al.[1] used non-uniform image patch model to solve low layer vision problems such as image de-noising and super resolution, as well as image restoration [2]. In addition, image segmentation also can be used for image filtering [3] .

Of course, three research topics: image segmentation, extraction of patches and annotations of patch content, can be realized through user interaction. For example, LabelMe [4] is an effective and convenient tool for internet users to make image annotations. However, faced with a large number of image data processing, the technical study of automatic image segmentation and patch structure extraction is necessary. A good method for image patch extraction can not only keep the original information structure, but also is consistent with human's subjective feeling, which is a big challenge. We in this paper present a semi-automatic algorithm for the extraction of multiple resolution patches, under the constraints of the large spatial scale of texture elements, sharpness and the number of iteration of the algorithm. The image patch extraction experiment results show that the new method is effective and feasible.

2 Related work

According to the complexity, images are classified into two categories: texture image and non-texture one. Due to the simple structure and content of non-texture images, some traditional image segmentation algorithms work well. However, texture images, natural landscape or humanistic scenic sites for example, are very common in daily life and contain complex texture. Segmentation for such images becomes more challenging than non-texture images. From the view of the segmentation results, the existing methods for texture image segmentation, can be classified into two categories: static image segmentation; extraction of multiple resolution patches with annotation. The former refers to the optimal segmentation results of a specific object from the image, and the result is the only one; the latter refers to sequential segmentation results from one image according to certain rules.

Static image segmentation includes methods based on gray threshold, region, edge detection or a combination with some new theories. Image segmentation theory emphasized [5] that the segmentation should be consistent with human cognition, and imitate human perceptual cognition to achieve. Moreover, the theory recognized the important position of lines in the image and used the tree structure to describe the image content and the relationship between the various patches. In recent years, many image segmentation methods successfully combined with some specific theories or techniques. For example, on the basis of lazy learning and Markov random field, Tighe J. et al. [6] put forward a non-parametric method to analyze images and annotated image region category. Li L. et al. [7] presented an automatic method for classification, segmentation and annotations by constructing a hierarchical model with image patches, objects and scenes such three levels. Literature [8] combined region information from the bottom to the top and local information descriptor from the top to the bottom to realize semantic segmentation and object annotation.

Combined with machine learning or prior knowledge, image patch segmentation can be more effective. For instance, literature [9] presented an image annotation method using sparse representation to prevent the over fitting and outliers in semi-supervised learning. In 2014, literature [10] gave a machine learning approach based on prior knowledge and texture features—Ratsnake, achieving ideal effect for a variety of medical image segmentation. Literature [11] used image depth information and image prior information to detect objects and segment images. And literature [12] annotated image objects in a pixel level by combining the regions-based with sample detectors-based segmentation algorithm. Xu L. et al. [13] extracted the patch structure from images by using relative total variation theory. This method was able to eliminate some details or textures in images, and keep the image edges, which can be applied to the edge extraction and image simplification.

Multiple resolution patches can be extracted by using threshold-based method and regional growth-based one. Li C. et al. segmented images by combining the multi-resolution technique and Gibbs sampling technology together[14]. Kim J. et al.[15] applied the pyramid representation of images and the wavelet transform to generate the multiple resolution patches, thus realized an efficient image segmentation integrating with watershed segmentation algorithm. Obrador P. [16] proposed the pyramid design strategy to extract multiple resolution patches on the grounds of maximum likelihood principle, while preserving the edge and some detail regions. Since the simple multiple resolution technology may lead to objects' inaccurate geometry information in segmentation results, an effective improvement is to combine with machine learning methods [17]. Yin S. et al. [18] presented a multilevel image segmentation method which is through fuzzy c-segmentation entropy maximization and graph cut optimization. In 2015, a semantic segmentation method which uses context information to construct a deep neural network was presented [19].

From those literatures referred above, our method in this paper falls into the multiple resolution segmentation group.

3 Extraction of multiple resolution patches

Our algorithm is based on the relative total variation patch extraction(rTVPE)[13]. rTVPE generated the optimum extraction results of patch structure by adjusting parameters: smoothness α, texture spatial scale β, sharpness γ and algorithm iteration number η. After testing the effects of these parameters in structure extraction, we find that when the smoothness α changes, the segmented image patches change significantly. This inspires us to present a method of extracting multiple resolution patches from images by setting this parameter as sequence values and obtaining a series of output segmented images shown as different resolution patches. The main frame of our algorithm is shown in Fig.1, and the iteration stop criterion and iteration step length are discussed in Section 3.1 and 3.2..

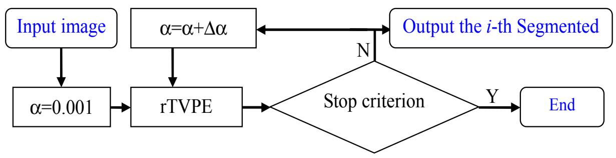

Fig 1. Algorithm flow.

3.1 Parameter analysis

Sharpness γ in the [13] is a real number, and typically varies in a small range [0.001,0.03]. Through a lot of experiments, we find that the patch structure extracted from an image shows clear edge if γ = 0.001. Therefore, we use the value in subsequent experiments.

The texture spatial scale parameter β mainly controls the size of the window for computing the windowed variations. It depends on the scale of the texture elements, therefore plays an important part in the separation of the texture structure. As for different images, β values to the interval (0,8]. The texture-suppression effect can be obtained by increasing β. We let β=3 to get a moderate effect in practice.

As for iterations, we also find that when the value of iterations η is adjusted and the other parameters are fixed, extraction results illustrate that η =3,4 or 5 is enough for patch structure extraction. So we set the number of iterations for the default value, η =4.

During the experiment, the most significant among those parameters is smoothness α. As its value becomes bigger, the patches in images would transit smoother, thus strengthening the fuzzy degree. When $\alpha > 2$, the image is often merged into one patch; By contrast, while α is set to be small, images still maintain the original complexity with inconspicuous characteristics of image patch structure. Therefore, the value of α too large or too small would eventually make the extraction of multiple resolution patches from image meaningless. Based on these tests, we set α in the range [0.001,2]. Moreover, the number of patches would change a lot when varied from [0.001, 0.1], but not changes too much when $\alpha \in [0.1,2]$. Therefore, we set the smoothness parameter with non-uniform sequence value with formula (1).

$$\alpha_0 = 0.001, \quad \alpha_i = \alpha_{i-1} + \Delta\alpha_i, i = 1,2,...,n \quad (1)$$

The calculation of the sequential increment is a piecewise function:

$$\Delta\alpha_i = \begin{cases} 0.099/n_0, & 1 \le i \le n_0 \\ 1.9/(n-n_0), & n_0 < i \le n \end{cases} \quad (2)$$

where, $n_0 = \lfloor 0.8n \rfloor$ and $\lfloor x \rfloor$ rounds the elements of x to the nearest integer towards minus infinity. In this way, by adjusting the parameter α from 0.001 to 2 for an input image, a series of images with different resolution patches are obtained.

3.2 Iteration stop criterion

When modifying rTVPE 13] to automatically extract a series of images with different resolution patches by setting the sequential value of the smoothness α, we need to construct an iteration stop criterion. During the iterative process, the differences between two consecutive image patches in the segmented image sequence are the number of patches and the amount of the structure information from the original image. Based on the differences, the optimal stop conditions of iterative algorithm can be defined.

We hope that the image segmentation will contain original structure information as much as possible, that is, control the image information in what segmentation we choose accounted for a large proportion in the original image. According to this, we can list the first objective function:

$$\max \quad f_1 = \frac{1}{w \cdot h} \sum_{i=1}^{m} p_i \quad (3)$$

In formula (3), $m = \min\{\tau, k\}$ means the top m patches of containing most number of pixels after image segmentation. τ set by users in the experiment is the maximum number of patches that should be calculated. In our experiments, we let τ be a constant value, τ=100. p_i is the number of pixels included in the i-th patch after segmentation. w and h are the width and height of the original image. k is the actual number of the current image segmentation. When α values smaller, k becomes bigger.

In addition, the optimal number of patches among similar images often obeys certain statistical laws. Here we assume that the optimal number of patches μ in segmentation obeys normal distribution $N(\mu, \sigma^2)$, and the parameters μ and σ^2 can be calculated through the annotation database, or given through users' investigation. So the second objective function of the optimal segmentation is the maximization of a Gauss function

$$\max \quad f_2 = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(m-\mu)^2}{2\sigma^2}} \quad (4)$$

From the formula (3) and (4), we are dealing with a multi-objective decision making problem. The weighted objective function we established is:

$$\max \quad z = w_1 f_1 + w_2 f_2 \quad (5)$$

where w_i (i=1,2) is the weight of two objective functions determined by users according to the need of the segmentation.

Base on the formula (5), the algorithm iterative stop criterion is $z_i < z_{i-1}$ or $|z_i - z_{i-1}| < \delta$, that is, the iteration stops when the objective function becomes smaller or the objective function value increases relatively slowly, or both.

In addition, during the experiment, we found that some of the noise contained by images would interfere with the segmentation. So we use the image expansion to deal with the original image in advance, and then extract the multiple resolution patches from images.

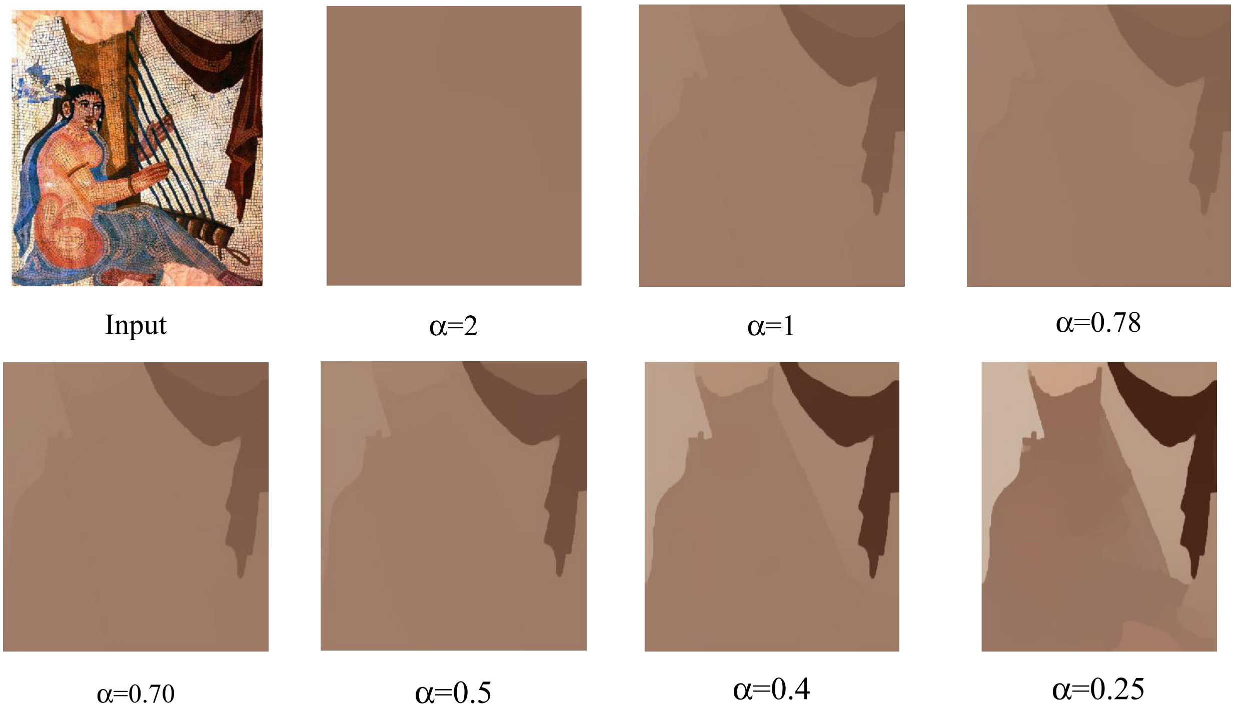

Fig 2. A sequence of multiple resolution extraction for one image of art

4 Experiment

In order to verify the feasibility and validity of the proposed method, these experiments partly extract different resolution patches in different types of images. These experiments are taken in PC with Core(TM)i7-4700MQ CPU@2.4GHz, 8G@RAM, and use Matlab as programming language. During the experiment, we take statistical sampling for image annotations from the SUN database[20], and use the method of maximum likelihood to estimate parameters in the formula (4). The results of parameters estimation are μ=4.18 and $\sigma^2=0.907^2$.

4.1 Patch sequence

The first experiment is to illustrate the extraction algorithm of multiple resolution patches with an image [13]. Through automaticlly adjusting the image smoothness parameter α, a series of segmented images shown as different resolution patches are obtained. Figure 2 shows the experiment results with a series of α values. When α value changes from big to small, the number of patches increases correspondingly.

4.2 Comparison

Comparing with other methods, the experiment is mainly taken in the view of the optimal segmentation result, as shown in Figure 3. Although our method automatically stops in the optimal segmentation, its effect is not worse than that from SUN database[20] which is an example from LabelMe, and the rTVPE [13]which is obtained by carefully setting threshold of several parameters.

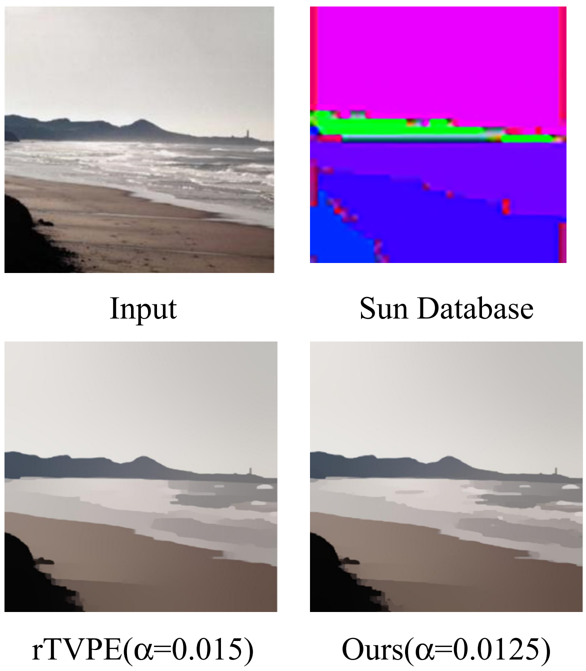

Fig 3. Comparison of the optimal segmentation results of different methods

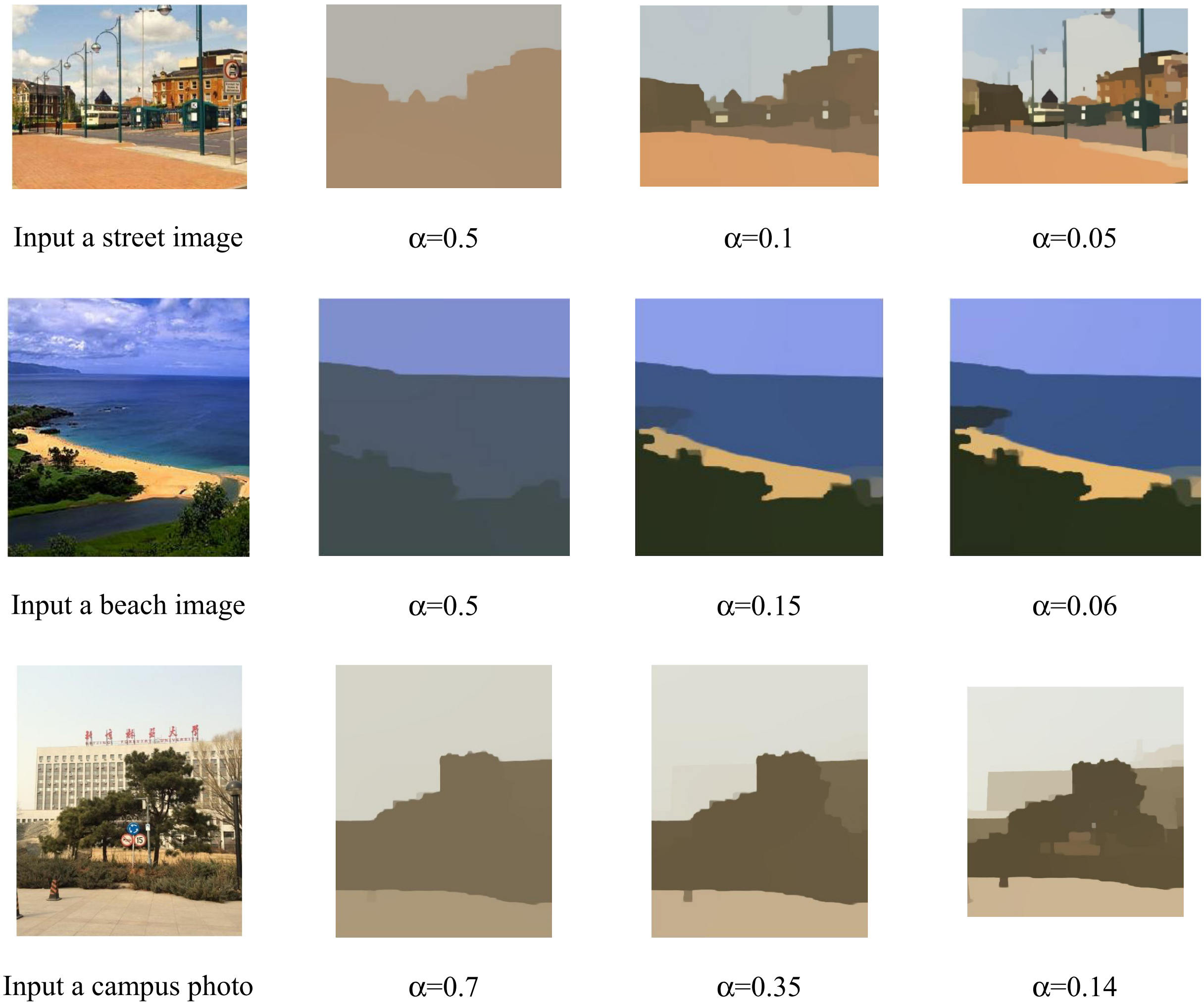

Fig 4. More experiment results.

4.3 More test

More images are tested and the extracted patches are listed in Figure 4. In this figure, both of the input street image and the beach one come from SUN database[20]; and the campus photo is taken by ourselves. All these input images are separated by the multiple resolution patch extraction algorithm. From Figure 4, we can find out that the results of segmented images, in general, reflect the structure of the original images in different scale.

5 Conclusion

Based on the analysis of rTVPE algorithm [13] and its parameters: smoothness, texture spatial scale, sharpness, iteration number and by setting the last three parameters, we successfully construct an semi-automatic algorithm for extracting the multiple resolution patch sequence. It is an iterative algorithm by dynamically adjusting the image smoothing parameter. In addition, we present a piecewise function for determining the iteration step length and a double-objective optimization criterion to stop the iterative algorithm. Experiment results show the effectiveness of the proposed method.

This algorithm also remains some limitations and problems for further study. One of them is that during the experiments, the iterative stop condition is obtained through the analysis of image annotations, which limits the more widely application of the algorithm, because annotated images are now still rare. Therefore, how to build the optimal stop condition of the iteration without annotations needs to do more deep research.

Acknowledgements

This research work is supported by National Natural Science Foundation of China with project Nos. 61372190, 61571046, and National Undergraduate Innovation Project No. G201410022044.

References:

[1] De Smet V, Namboodiri V P, Van Gool L. Nonuniform Image Patch Exemplars for Low Level Vision[J]. *IEEE Workshop on Applications of Computer Vision*, 2013:23 - 30.

[2] Mobahi H, Rao S R, Ma Y. Data-Driven Image Completion By Image Patch Subspaces[J]. *Picture Coding Symposium Pcs*, 2009:1 - 4.

[3] Guo G, Dyer C R. Patch-based Image Correlation with Rapid Filtering[C]. //*IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2007:1-6.

[4] Torralba A, Russell B C, Yuen J. LabelMe: Online Image Annotation and Applications[J]. *Proceedings of the IEEE*, 2010, 98(8):1467 - 1484.

[5] Wardhani A, Gonzalez R. Image Structure Analysis for CBIR[C]. //*Proc Digital Image Computing Techniques & Applications Dicta*. 1999:166.

[6] Tighe J, Lazebnik S. SuperParsing: Scalable Nonparametric Image Parsing with Superpixels[J]. *Lecture Notes in Computer Science*, 2010，pp.352-365.

[7] Li L, Socher R, Fei-fei L, Towards total scene understanding: Classification, annotation and segmentation in an automatic framework[C]. //*IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009:2036 - 2043.

[8] Arbelaez P, Hariharan B, Gu C, Gupta S, Bourdev L, Malik J. Semantic segmentation using regions and parts[C]. //*IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012:3378 - 3385.

[9] Nakashika T, Okumura T, Takiguchi T, Ariki Y. Sparse representation for outliers suppression in semi-supervised image annotation[C]. //*IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013:2080 - 2083.

[10] Iakovidis D K, Goudas T, Smailis C, et al., Ratsnake: a versatile image annotation tool with application to computer-aided diagnosis.[J]. *Scientific world journal*, 2014, 5(1):149-168.

[11] Saurabh Gupta, Ross Girshick, Pablo Arbeláez, Jitendra Malik. Learning Rich Features from RGB-D Images for Object Detection and Segmentation. *Computer Vision-ECCV2014, Lecture Notes in Computer Science,* Volume 8695, 2014, pp 345-360.

[12] Tighe J, Lazebnik S. Finding Things: Image Parsing with Regions and Per-Exemplar Detectors[C]. //*IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2013:3001-3008.

[13] Xu L, Yan Q, Xia Y, Jia J. Structure extraction from texture via relative total variation[J]. *Acm Transactions on Graphics*, 2012, 31(6):439-445.

[14] Li C -. Multiresolution image segmentation integrating Gibbs sampler and region merging algorithm[J]. *Signal Processing*, 2003, 83:67–78.

[15] Kim J, Kim H. Multiresolution-based watersheds for efficient image segmentation[J]. *Pattern Recognition Letters*, 2003, 24:473–488.

[16] Obrador P. Multiresolution color patch extraction[J]. *Visual Communications & Image Processing*, 2006.

[17] Belgiu M, Drăguţ L. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery[J]. *Isprs Journal of Photogrammetry & Remote Sensing*, 2014, 96(4):67–75.

[18] Yin S, Zhao X, Wang W, Gong M. Efficient multilevel image segmentation through fuzzy entropy maximization and graph cut optimization[J]. *Pattern Recognition*, 2014, 47(9):2894–2907.

[19] Sharma A, Tuzel O, Jacobs D W. Deep Hierarchical Parsing for Semantic Segmentation[J]. *Eprint Arxiv*：1503, 2015.

[20] SUN database. *http://groups.csail.mit.edu/vision/SUN/*, 2015-2-5.