

Research and Application on Tangka Image Segmentation Algorithm

Yan Xu, WeiLan Wang

Northwest University for Nationalities Mathematics and computer science college, Lanzhou 730030,
China

dcr@126.com

Keywords: Tangka image, segmentation algorithm, image segmentation, clustering analysis, multi-feature combination.

Abstract. This paper studies the key problems of many Tangka image segmentation algorithms and computational complexity and hard to applies to large scale images segmentation. As a exploratory research of Tangka image, the paper fully took the comprehensive and robustness of algorithm into account. The image segmentation based on saliency area detection had nice translational, multi feature combination and rotational invariance, so we focus on image segmentation. First, design the algorithm based on clustering analysis theory and segments the fuzzy image, especially the fuzzy edge. Then we improve the Tangka segmentation algorithm in order to meet high precision of the application. Meanwhile, Tangka image segmentation offer convenience for research in future, and the final recognition result was satisfy us according to the experiment.

Introduction

Image segmentation is the basis and key technology of image processing, pattern recognition. The results of segmentation can directly affect the other steps Image bears fuzziness in nature, which make the segmentation more difficult [1]. Since the uncertainties are non-randomness, we cannot use classical mathematics theory to handle it.

Tangka image analysis is an important technology in data mining, and it has been widely used in areas such as statistics, image processing, medical diagnosis, information retrieval, biology and machine learning [2]. Obviously, the problem of image segmentation is coincident with the problem of Tangka image. Image segmentation plays a fundamental role in computer vision as a requisite step in such tasks as object detection, classification, and tracking. In computer vision, a lot of Tangka image algorithms are not proper to be applied if the image data set is too big. Because the data of digital image processing in computer vision is often great and the complexity of the image itself is unpredictable, especially for the treatment of the natural color image sequence which is very large, the processing of image segmentation often takes more time and lacks of precision or it is unrealistic. Therefore, how to realize the real-time automatic quick image segmentation is still extremely important and the problems are yet to be solved effectively.

Algorithms

In this paper, we use clustering algorithm to segment Tangka image, because the clustering theory is very popular in the field.

According to the problems of the affinity propagation(AP) clustering algorithm existing huge storage and computational complexity and hard to be used in image data real-time processing, a new color image segmentation algorithm is proposed based on mean shift(MS) and affinity propagation algorithm named MSAP [3]. The proposed method preprocesses an input image by MS algorithm. The numbers of segmented regions, instead of the numbers of image pixels, are considered as the input data scale of AP algorithm. The average of the color vectors in each region is calculated and considered as an input data point of AP algorithm. Distances between data points are regards as similarity measure index, and then the AP algorithm is applied to perform globally optimized clustering and segmentation based on similarity matrix.

Function form is following as:

$$K = \frac{C}{h^d h^p} k \left(\frac{\|x\|^2}{h} \right) \quad (1)$$

Where, C is a constant; d and p express the spatial dimension; h is radius.

In this paper, $p=3$ and $d=2$.

Hierarchical clustering(HC) algorithm can obtain good clustering results, but it needs large storage and computational complexity for large image processing. A new color image segmentation algorithm based on mean shift and hierarchical clustering algorithm named MSHC is presented in this paper. MSHC algorithm preprocesses an input image by MS algorithm to form segmented regions that preserve the desirable discontinuity characteristics of image [4]. The number of segmented regions, instead of the number of image pixels, is considered as the input data scale of HC algorithm. The proximity between each cluster is calculated to form the proximity matrix, and then ward algorithm is employed to obtain the final segmentation results. MSHC algorithm is employed on color image and medical image segmentation. Node i and k , the distance between them is d as following below.

$$d_{ik} = \|x(i) - x(k)\|^2 \quad (2)$$

$$\begin{aligned} SSE &= \sum_i \|X - u\|^2 + \sum_j \|X - u\|^2 \\ &= \sum_i \left\| x - u - \frac{ni(uj - ui)}{ni + nj} \right\|^2 + \sum_j \left\| x - u - \frac{ni(uj - ui)}{ni + nj} \right\|^2 \\ &= SSE_i + \frac{ninj^2}{(ni + nj)^2} \|ui - uj\|^2 + SSE_j + \frac{ni^2nj}{(ni + nj)^2} \|ui - uj\|^2 \\ &= SSE_i + SSE_j + \frac{ninj}{ni + nj} \|ui - uj\| \end{aligned} \quad (3)$$

Where, SSE express sum of square error, so the distance is:

$$d = \sqrt{\frac{ninj}{ni + nj}} \|ui - uj\| \quad (4)$$

In order to obtain good and robust clustering results, a new clustering algorithm called clustering by data competition is proposed for large image segmentation processing. The proposed algorithm the clustering exemplars and clustering members according to the data energy and determines the best exemplars according to the data competition [5]. Then the proposed algorithm is combined effectively with the mean shift algorithm for large scale color image segmentations. Furthermore, the proposed algorithm has high segmentation efficiency, and gets a better image segmentation performance as well. An model scheme is shown in Fig. 1.

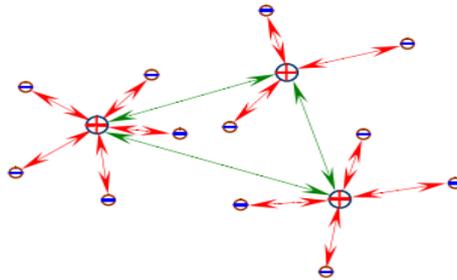


Fig. 1 Model scheme

Where the distance D express:

$$D = \begin{bmatrix} d_{11} & \dots & d_{1n} \\ \dots & \dots & \dots \\ d_{n1} & \dots & d_{nn} \end{bmatrix}$$

So the minimum distance is:

$$d_{\min} = \min \{d_{ij} | i = 1, 2, \dots, n; j = 1, 2, \dots, n; i \neq j\}$$

In order to solve the problem of the traditional spectral clustering existing huge storage and computational complexity and hard to be used in image data real-time processing, the cosine similarity is introduced into the process of image spectral segmentation. A fast image segmentation algorithm based on spectral clustering of Nystrom approximation and cosine similarity is presented. That the cosine similarity is calculated as the weight matrix of spectral clustering avoid to calculating the exponential operation and parameter setting and reduce the computational cost effectively.

Experiment

In the extraction of Tangka image information, the images are often only interested in certain parts. In order to analyze these goals of the images, they will be separated from the images, and on this basis it will be possible to further the goals of measurement and use of the image, that is, Tangka image segmentation technology which is one of the hot issues in recent years. An example of Tangka image is shown in Fig. 2.

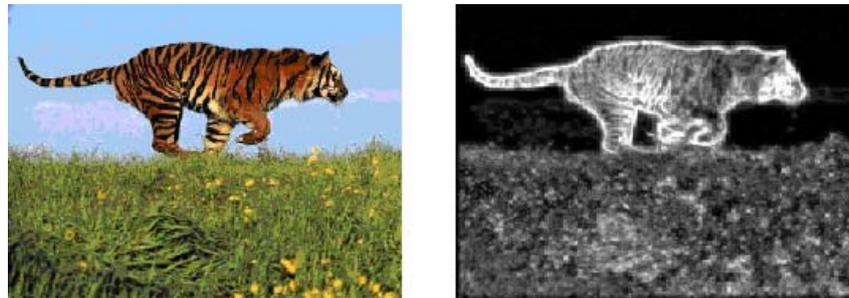


Fig. 2 Algorithm effect diagram

The experiment data is in Table 1.

Table 1 Datas illustration

1	4k2_far	400	4	2
2	Iris	150	3	4
3	Yeast	384	5	17
4	Wine	178	3	13
5	Segementation	210	7	19

In an experiment, a new method of the application on Tangka image segmentation, that is, the adaptive threshold segmentation algorithm based on the wavelet transform is proposed. The theory explain the algorithm in the image segmentation results and efficiency, and proves the algorithm is more obviously effective and practical. Application cases are as shown in Fig. 3 and Fig. 4.



Fig. 3 Thangka religious image core sample graph

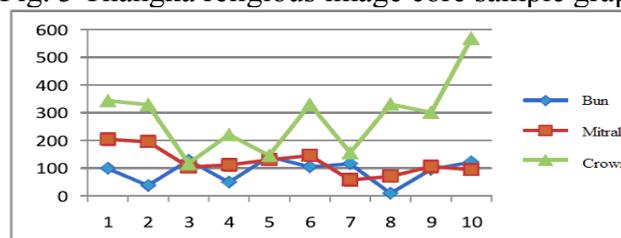


Fig. 4 Rate of color change feature (statistics parameter)

In order to solve the problem of the traditional image segmentation algorithm existing large storage and computational complexity, a new color image segmentation method combining twice used watershed and clustering theory of improved the weight matrix algorithm is presented in this paper. The image clustering uses the segmented regions, instead of the image pixels, the new method can effectively reduce the computational complexity of traditional clustering method by using secondary watershed algorithm. The new weight matrix also has certain self-adaptability.

$$g(x, y) = grad\{f(x, y)\} = \left\{ [f(x, y) - f(x-1, y)]^2 + [f(x, y) - f(x, y-1)]^2 \right\}^{0.5} \quad (5)$$

Where, $f(x, y)$ express the initial image, $grad$ express gradient operation.

According to image features and segmentation rules, the proposed algorithm defines image homogeneity, combines global information, local information, edge information, utilize global and local segmentation strategy. The algorithm calculates the initial cluster center by using histogram, measures the pixel indeterminacy by I , defines region merge rule based on I and region adjacency relationship.

Experimental results of gray segmentation prove that the proposed approach is effective than fuzzy logic methods. It can avoid the wrong segmentation phenomenon and get more homogeneous regions. In color segmentation, the independence of T , I , F and the comprehensive ability of using varied information are more obviously. How to select a suitable color space is a problem in color segmentation, different color space has its advantages and disadvantages and cannot replace each other.

Conclusions

The study can solve complex image segmentation problems with higher ability of uncertainty representation. It also solves the problem of information lacking using varied image information under this segmentation framework. My research provides new idea and method to solve the bottleneck of segmentation, and can be further extended to image enhancement, image de-noising, and other image processing areas, which has theoretical significance and wide application prospects.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No.60875006, No.61162021). Funds for the Central Universities 2015 (No. 31920150083).

References

- [1] Ying Liu, Dengsheng Zhang, Guojun Lu, Wei-Ying Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 2008, 40: 262-282.
- [2] Richard Szeliski. *Computer Vision: Algorithms and Application*. Springer, 2010.
- [3] Wong A K C, Sahoo P K. A gray-level threshold selection method based on maximum principle. *IEEE Transactions on System, Man, and Cybernetics*, 2009, 19(4): 866-871.
- [4] Abutaleb A S. Automatic thresholding of gray-level pictures using two-dimensional entropies. *Pattern Recognition*, 2012, 47(1): 22-32.
- [5] Cohen P R, Schrag R, Jones E, et al. The DARPA High Performance Knowledge Bases Project. *Artificial intelligence*, 2013, 19(4):25-49.