Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 55 (2015) 1015 – 1022

Procedia Computer Science

Information Technology and Quantitative Management (ITQM 2015)

Image Segmentation via Improving Clustering Algorithms with Density and Distance

Zhensong Chen[a,c,e], Zhiquan Qi[b,c], Fan Meng[a,c], Limeng Cui[c,d], Yong Shi[*,b,c]

[a]School of Management, University of Chinese Academy of Sciences, Beijing, 100190, China
[b]Research Center on Fictitious Economy and Data Science, CAS, Beijing, 100190, China
[c]Key Laboratory of Big Data Mining and Knowledge Management, CAS, Beijing, 100190, China
[d]School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing, 100190, China
[e]Sino-Danish Center for Education and Research, Beijing, 100190, China

Abstract

Image segmentation problem is a fundamental task and process in computer vision and image processing applications. It is well known that the performance of image segmentation is mainly influenced by two factors: the segmentation approaches and the feature presentation. As for image segmentation methods, clustering algorithm is one of the most popular approaches. However, most current clustering-based segmentation methods exist some problems, such as the number of regions of image have to be given prior, the different initial cluster centers will produce different segmentation results and so on. In this paper, we present a novel image segmentation approach based on DP clustering algorithm. Compared with the current methods, our method has several improved advantages as follows: 1) This algorithm could directly give the cluster number of the image based on the decision graph; 2) The cluster centers could be identified correctly; 3) We could simply achieve the hierarchical segmentation according to the applications requirement. A lot of experiments demonstrate the validity of this novel segmentation algorithm.

Peer-review under responsibility of the Organizing Committee of ITQM 2015

Keywords: Image Segmentation, Clustering Algorithm, Feature Representation;

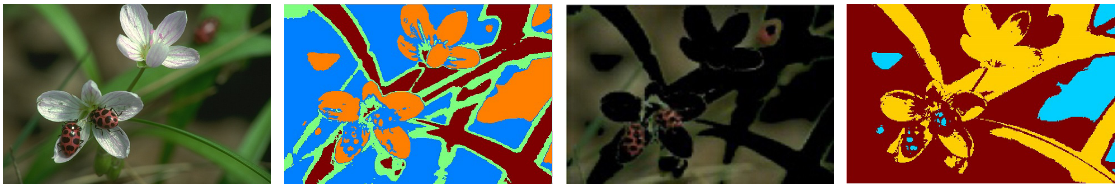
1. Introduction

Image segmentation is a fundamental task in computer vision and image processing applications. As we all know, image segmentation refers to a procedure of dividing the input image into several regions according to the visual characteristics shared by the pixels. In the past several decades, a lot of segmentation algorithms have been proposed. And these segmentation methods could be divided roughly into the following four categories: thresholding, clustering, edge or contour detection and region extraction [1].

Clustering method is one of the most popular algorithm in image segmentation domain. It is an approach of classifying patterns or data into categories, which is on the basis of the samples in the same group have the higher similarity than the ones in different groups. These methods project the input image into their features spaces

* Corresponding author. Tel.: +86-010-82680698.
E-mail address: yshi@ucas.ac.cn.

Peer-review under responsibility of the Organizing Committee of ITQM 2015
doi:10.1016/j.procs.2015.07.096



(a) Original (b) K-means (c) FCM (d) Proposed

Fig. 1. An illustrative comparison on the segmentation performance of the original image (a), using different algorithms: (b) K-means, (c) Fuzzy c-means, (d) our proposed algorithm.

firstly. And then the segmentation of the images is often obtained by applying clustering algorithms on feature spaces along with giving the cluster numbers by human.

Particularly, K-means clustering algorithms [2, 3] and fuzzy c-means clustering algorithms [4–6] are the two popular examples of the state-of-the-art clustering methods. These methods are able to generate a partition of images under some conditions including giving the cluster number prior. However, the performance of segmentation is sensitive to cluster centers and numbers, which are really difficult to be identified directly and prior by human beings.

In this work, we use an improving clustering algorithm, which could determine the cluster numbers and cluster centers based on the decision graph [7], on the representation of the input image in color feature spaces. An illustrative example is given in Figure 1. It demonstrates that compared with K-means and fuzzy c-means clustering, our proposed method could yield the image segmentation based on the objects of input image.

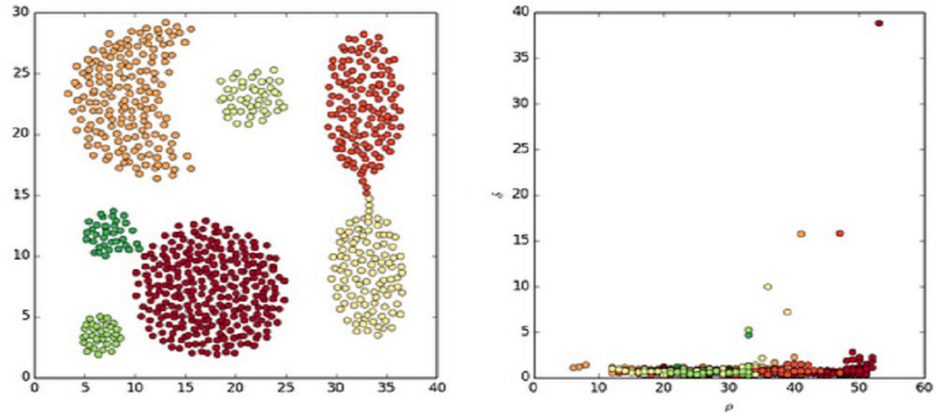
The rest of this paper is organized as follows. Section 2 will introduce the related work proposed in recent years. We will present the clustering approach and features we used in this paper and describe our posed image segmentation algorithm in section 3. While Section 4 will show the experiments to evaluate our algorithm. Finally, conclusions will be made in Section 5.

2. Related Work

In the last decades, there are a lot of image segmentation approaches haven been proposed and developed. Since we are focus on the clustering-based segmentation algorithms in this work, we will review the related work along this direction in details at the following part. For other methods, a good review can be found in [8].

K-means clustering algorithm is an iterative technique used to classify the data points into K groups according to the similarity between them. It is developed by J. Macqueen [9] and is one of the simplest unsupervised learning algorithms that solve the well known clustering. After then, many researchers in the image processing domain applied it on image segmentation to improve the performances. Oliver et. al. [2] proposed a new strategy for clustering segmentation which uses K-means clustering integrating region and boundary information. Jumb et al. [3] integrated K-means and thresholding techniques to implement the segmentation of color image. Kochra [10] used the Hill-climbing with K-means algorithm for color image segmentation. An example of the segmentation of K-means clustering algorithms on color image could be seen in Figure 1.

Fuzzy c-means (FCM) clustering is another popular clustering algorithm that most used in image segmentation problems. FCM techniques introduces the fuzzy concept into image segmentation problems so that an object can belong to several classes as same time. It is an unsupervised technique and it basic idea is that clustering the data points by iteratively minimizing the cost function, which is dependent on the distance between the pixels and the cluster center in the feature space. Han and Shi [11] proposed a fuzzy ant system pixel clustering method, which extracting three kinds of features including grayscale value, gradient value and pixel neighborhood information, for color image segmentation. Tan [12] presented the Region Splitting and Merging-Fuzzy C-means Hybrid Algorithm (RFHA), which is an adaptive unsupervised clusterin approach for color image segmentation, together with histogram thresholding techniques. Liew et al. [13] developed an adaptive spatial fuzzy c-means clustering algorithm for the segmentation of three-dimensional (3-D) magnetic resonance (MR) images. An example of the segmentation of FCM clustering algorithms on color image could be seen in Figure 1.



(a) Point Distribution (b) Decision Graph

Fig. 2. The algorithm in two dimensions. (a) Point distribution. (b) Decision graph for the data in (a). Different colors correspond to different clusters.

Furthermore, there are many clustering algorithms including spectral clustering [14], mean shift clustering [15] and so on are used in the image segmentation domain. These clustering-based segmentation approaches are often applied along with other techniques including automatic scale selection technique, histogram thresholding techniques and external information such as texture information and spatial information. Especially, support vector machine (SVM) [16–21] is another potential technique which could be incorporated with clustering algorithms. Besides, some researchers tried to use two different clustering-based approaches simultaneously: one algorithm is used to over segment the image and the other one to merger the little regions.

Although these clustering-based segmentation have solved the image segmentation problem to some extent, they are not able to segment the image without giving the clustering number prior. In addition, these performances are sensitive to the selection of initial cluster centers.

3. Proposed Approach

3.1. Clustering Methods

According to the studies on clustering algorithms, R. Alex and L. Alessandro [7] proposed a novel clustering approach which could be called DP clustering method. This method could clustering data by fast search and find of density peaks. And it is proposed under the basic assumptions: cluster centers are often surrounded by points who has lower density and have a relatively large distance from these point with higher density. This method could identity the cluster centers and numbers based on the decision graph (as shown in Figure 2) without any other priori knowledge.

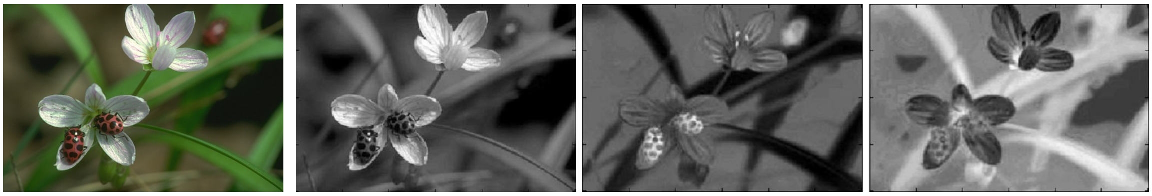
Based on this basic principles, we could compute the two quantities for each point i: its density ρ_i and its distance δ_i away from the points with higher density. The definition of density ρ_i of point i is shown as follows:

$$\rho_i = \sum_j \exp^{-\frac{d_{ij}^2}{d_c^2}} \tag{1}$$

where d_{ij} denoted the distance between data point i and point j, d_c is a cutoff distance. Obviously, ρ_i could denote the points distribution around the point i.

We could measure the distance δ_i through computing the minimum distance between point i and the points whose density is higher:

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}) \tag{2}$$



(a) Original (b) Color Channel Features

Fig. 3. The three color channels features of the original image (a) are shown in (b).

For the point with the highest density, we take $\delta_i = \max_j(d_{ij})$. And then we can get the decision graph composed with the density ρ and distance δ to show the plot of δ_i as a function of ρ_i for each point (see the right picture in Figure 2). Hence, as anticipated from this decision graph, the points which has high distance (δ) along with high density (ρ) are the cluster centers.

After we select the cluster centers based on the decision graph, the remaining points that are not the cluster centers should be assigned to the same cluster as its nearest neighbors of higher density after the cluster centers are found. The cluster assignment is performed in a single step compared with other clustering algorithms where an objective function is optimized iteratively [22]. An illustration of this clustering algorithm in two dimensions data is shown in Figure 2.

3.2. Image Segmentation using DP Clustering Algorithms

For an input image, the first step of clustering based segmentation approaches is projecting the image into the feature spaces. In our work, we will choose the color channels as basic features to representant the image (as shown in Figure 3).

Based on the above processing of the original image, we could get the representation of the input image on color channel features. After then, we apply the DP clustering algorithm which we described in section 3.1 on these representations. Our proposed algorithm requires computing the distances between all the input data, and we take the Euclidean distance as the main measurement of data distances in this computation. And we take the Gaussian kernel as the basic measurement of density ρ_i of the data point i in our work. We could get the density ρ_i from equation (1). And then we could compute the distance δ_i for each point i according to the follow equation (3):

$$\delta_i = \begin{cases} \min\limits_j(d_{ij}) & \rho_j > \rho_i \\ \max\limits_j(d_{ij}) & \rho_i \text{ is the highest density} \end{cases} \tag{3}$$

Based on the results computed by equation (1) and (3), we can compose the decision graph which could be used to identify the cluster centers and number, and eventually achieving the image segmentation.

Together with the description of projecting the input image into color spaces and the basic idea of DP clustering algorithm, the novel segmentation approach could be described as follows:

(1) Transforming the input image into feature representation.

- Read the original image data to obtain the representation in three color channels, as shown in Figure 3.

(2) Identifying the cluster centers and number.

- Computing the density ρ and distance δ by using the equation (1) and (3) respectively. And then composing the decision graph based on the density and distance, as shown in Figure 4.
- According to the rules presented before, choosing the data points with high density (ρ) and large distance (δ) as the cluster centers. And then, we can obviously figure out the cluster number.

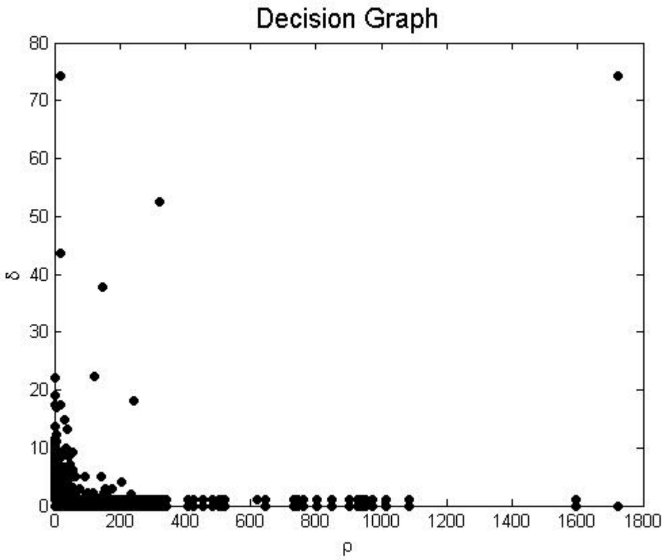


Fig. 4. The decision graph composed with the density ρ and distance δ.

(3) Assigning the remaining points to the clusters.

- Marking point x_i with the same label of point x_j if satisfying the follow two conditions: (1) $\rho_j > \rho_i$ and (2) $d_{ij} = \min_{l \neq i}(d_{il})$.

(4) Achieving the final segmentation based on the labels marked through last step, as shown in Figure 5.

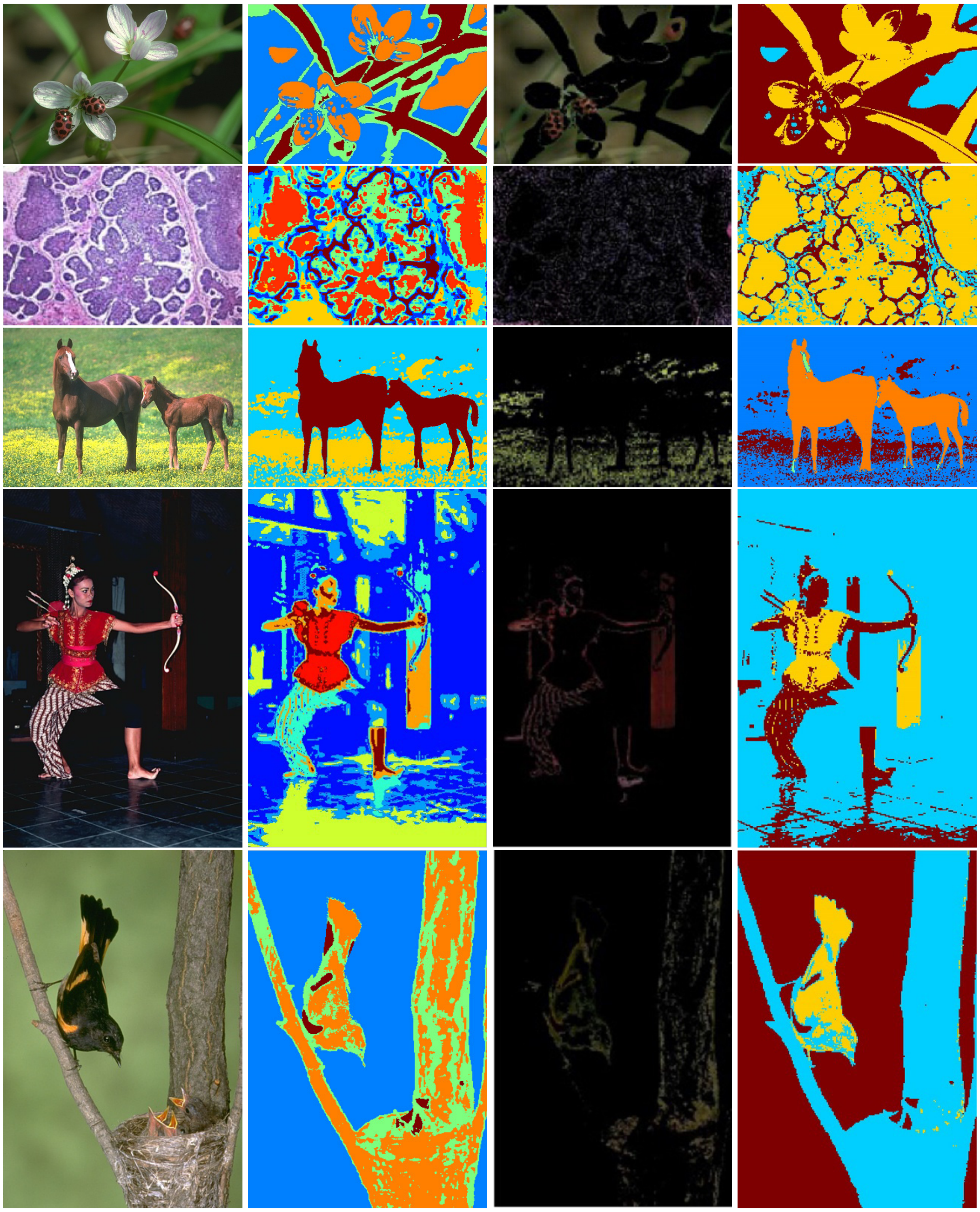
(a) Original Image (b) Segmentation Result

Fig. 5. The segmentation result (b) achieved by our proposed cluster-based image segmentation algorithm on the original image (a).

4. Experiments

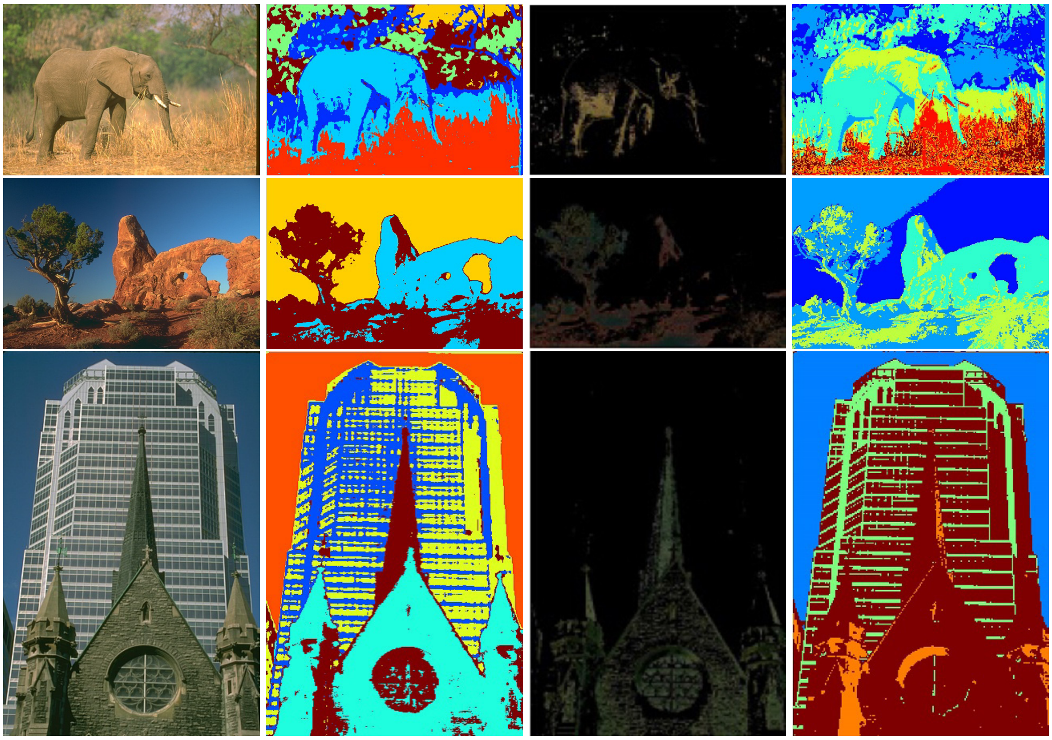
According to the algorithm description, the value of d_c is the only parameter of this novel clustering-based image segmentation approach that should be given prior. Varying d_c for the same input image will produce mutually consistent results. As a rule of thumb, one can choose d_c as a certain percentage of the maximum value of d_{ij}. In our experiment, we take $d_c = (0.5\%)d_{max}$ ($d_{max} = \max_{i,j}(d_{ij})$) as the experience value.

We have applied the proposed algorithm on Berkeley image database (BSDS500) [23]. And the state-of-art clustering-based algorithms, such as K-means, Fuzzy C-Means (FCM) [24] have been considered to compare with our new proposed segmentation approach, as shown in Figure 6 and Figure 7.



(a) Original (b) K-means (c) FCM (d) Proposed

Fig. 6. Comparison of different clustering-based image segmentation approaches. From left to right: (a) Original Image, (b) K-means based Segmentation, (c) FCM based Segmentation, (d) Proposed Method.



(a) Original (b) K-means (c) FCM (d) Proposed

Fig. 7. Comparison of different clustering-based image segmentation approaches. From left to right: (a) Original Image, (b) K-means based Segmentation, (c) FCM based Segmentation, (d) Proposed Method.

In contrast with image classification, image segmentation performance evaluation remains subjective. Borra et al. [25] pointed out that the segmentation or grouping performance can be evaluated only in the context of a task such as object recognition. This is because it is really difficult to identity what is a correct segmentation. And the results are good or not depends a lot on the purposes of image processing. But, our novel clustering-based segmentation approach could freely and directly select the cluster numbers based on the decision graph according to the requirement of the tasks. And we also could simply achieve the hierarchical segmentation if the applications need.

5. Conclusion

In this paper, we have proposed a novel image segmentation approach based on the DP clustering algorithm. This segmentation method could determine the cluster number and centers directly based on the decision graph, which is composed with the density ρ and distance δ. And the hierarchical segmentation could also be easily achieved via our segmentation approach. Extensive experimental results show that the proposed approach is a good compromise in-between state-of-art methods. In conclusion, our proposed method could be a feasible preprocessing method for operations such as pattern recognition and image semantic annotation. In future work, we plan to explore how to select the vale of parameter d_c automatically based on the input image.

Acknowledgements

The authors would like to express their sincere thanks to the associate editor and the reviewers who made great contributions to the improvement of this paper. This work was partially supported by the major project of NSFC

(Grant No. 71331005), the international (regional) cooperation project of NSFC (Grant No. 71110107026) and the grants from National Natural Science Foundation of China (Grant No. 61402429). .

References

[1] N. R. Pal, S. K. Pal, A review on image segmentation techniques, Pattern recognition 26 (9) (1993) 1277–1294.

[2] A. Oliver, X. Munoz, J. Batlle, L. Pacheco, J. Freixenet, Improving clustering algorithms for image segmentation using contour and region information, in: Automation, Quality and Testing, Robotics, 2006 IEEE International Conference on, Vol. 2, IEEE, 2006, pp. 315–320.

[3] V. Jumb, M. Sohani, A. Shrivas, Color image segmentation using k-means clustering and otsus adaptive thresholding, Int. J. Innov. Technol. Explor. Eng 3 (9) (2014) 72–76.

[4] B.-Y. Kang, D.-W. Kim, Q. Li, Spatial homogeneity-based fuzzy c-means algorithm for image segmentation, in: Fuzzy Systems and Knowledge Discovery, Springer, 2005, pp. 462–469.

[5] Z. Ji, Y. Xia, Q. Chen, Q. Sun, D. Xia, D. D. Feng, Fuzzy c-means clustering with weighted image patch for image segmentation, Applied soft computing 12 (6) (2012) 1659–1667.

[6] K. S. Tan, N. A. M. Isa, W. H. Lim, Color image segmentation using adaptive unsupervised clustering approach, Applied Soft Computing 13 (4) (2013) 2017–2036.

[7] A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks, Science 344 (6191) (2014) 1492–1496.

[8] Z. Ma, J. M. R. Tavares, R. N. Jorge, T. Mascarenhas, A review of algorithms for medical image segmentation and their applications to the female pelvic cavity, Computer Methods in Biomechanics and Biomedical Engineering 13 (2) (2010) 235–246.

[9] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1, Oakland, CA, USA., 1967, pp. 281–297.

[10] S. Kochra, S. Joshi, Study on hill-climbing algorithm for image segmentation.

[11] Y. Han, P. Shi, An improved ant colony algorithm for fuzzy clustering in image segmentation, Neurocomputing 70 (4) (2007) 665–671.

[12] K. S. Tan, N. A. M. Isa, W. H. Lim, Color image segmentation using adaptive unsupervised clustering approach, Applied Soft Computing 13 (4) (2013) 2017–2036.

[13] A.-C. Liew, H. Yan, An adaptive spatial fuzzy clustering algorithm for 3-d mr image segmentation, Medical Imaging, IEEE Transactions on 22 (9) (2003) 1063–1075.

[14] C. J. Tilton, Image segmentation by region growing and spectral clustering with natural convergence criterion, in: International geoscience and remote sensing symposium, Vol. 4, INSTITUTE OF ELECTRICAL & ELECTRONICSENGINEERS, INC (IEE), 1998, pp. 1766–1768.

[15] W. Tao, H. Jin, Y. Zhang, Color image segmentation based on mean shift and normalized cuts, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 37 (5) (2007) 1382–1389.

[16] V. N. Vapnik, V. Vapnik, Statistical learning theory, Vol. 1, Wiley New York, 1998.

[17] Z. Qi, Y. Tian, Y. Shi, Robust twin support vector machine for pattern classification, Pattern Recognition 46 (1) (2013) 305–316.

[18] Z. Qi, Y. Xu, L. Wang, Y. Song, Online multiple instance boosting for object detection, Neurocomputing 74 (10) (2011) 1769–1775.

[19] Y. Tian, Z. Qi, X. Ju, Y. Shi, X. Liu, Nonparallel support vector machines for pattern classification, Cybernetics, IEEE Transactions on 44 (7) (2014) 1067–1079.

[20] Z. Qi, Y. Tian, Y. Shi, Successive overrelaxation for laplacian support vector machine.

[21] Z. Qi, Y. Tian, Y. Shi, Efficient railway tracks detection and turnouts recognition method using hog features, Neural Computing and Applications 23 (1) (2013) 245–254.

[22] G. McLachlan, T. Krishnan, The EM algorithm and extensions, Vol. 382, John Wiley & Sons, 2007.

[23] D. Martin, C. Fowlkes, The berkeley segmentation database and benchmark, Computer Science Department, Berkeley University. http://www. eecs. berkeley. edu/Research/Projects/CS/vision/bsds.

[24] M. H. Rahman, M. R. Islam, Segmentation of color image using adaptive thresholding and masking with watershed algorithm, in: Informatics, Electronics & Vision (ICIEV), 2013 International Conference on, IEEE, 2013, pp. 1–6.

[25] S. Borra, S. Sarkar, A framework for performance characterization of intermediate-level grouping modules, Pattern Analysis and Machine Intelligence, IEEE Transactions on 19 (11) (1997) 1306–1312.