Neural Comput & Applic (2016) 27:1445–1450
DOI 10.1007/s00521-016-2287-7

ORIGINAL ARTICLE

Histogram-based automatic segmentation of images

Enver Küçükkülahlı[1] · Pakize Erdoğmuş[2] · Kemal Polat[3]

Received: 28 January 2016 / Accepted: 28 March 2016 / Published online: 11 April 2016

Abstract The segmentation process is defined by separating the objects as clustering in the images. The most used method in the segmentation is *k*-means clustering algorithm. *k*-means clustering algorithm needs the number of clusters, the initial central points of clusters as well as the image information. However, there is no preliminary information about the number of clusters in real-life problems. The parameters defined by the user in the segmentation algorithms affect the results of segmentation process. In this study, a general approach performing segmentation without requiring any parameters has been developed. The optimum cluster number has been obtained searching the histogram both vertically and horizontally and recording the local and global maximum values. The quite nearly values have been omitted, since the near local peaks are nearly the same objects. Segmentation processes have been performed with *k*-means clustering giving the possible centroids of the clusters and the optimum cluster number obtained from the histogram. Finally, thanks to histogram method, the number of clusters of *k*-means clustering has been automatically found for each image dataset. And also, the histogram-based finding of the number of clusters in datasets could be used prior to clustering algorithm for other signal or image-based datasets. These results have shown that the proposed hybrid method based on histogram and *k*-means clustering method has obtained very promising results in the image segmentation problems.

Keywords Histogram · Segmentation · Clustering · Image processing

✉ Kemal Polat
kpolat@ibu.edu.tr

1 Department of Computer Technologies, Duzce Vocational School, Duzce University, 81010 Duzce, Turkey

2 Department of Computer Engineering, Faculty of Engineering, DuzceUniversity, 81620 Duzce, Turkey

3 Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Abant Izzet Baysal University, 14280 Bolu, Turkey

1 Introduction

The process of segmentation can be defined as the separation of objects within image [1]. In a general sense, segmentation is used in order to find limits or objects within image. At the end of the process, it would be easier to analyze and define the image.

Segmentation has been often used in many areas such as image compression, arrangement, processing medical image, object determination in satellite images, facial recognition, finger print recognition and traffic control systems.

One of the most common methods used in segmentation process is *k*-means clustering algorithm. Clustering is one of the most popular data analysis and data mining techniques. The most popular technique for clustering is based on *k*-means divided into *k*-clustering of database [2]. *k*-means algorithm was developed by Macqueen in 1967 [3]. It is a method based on the idea that center of gravity of cluster elements represent the set [4]. *k*-means algorithm divided datasets input by the user and composed of n data into k pieces of cluster. Cluster relation is determined by the distance of objects to cluster centers.

In order for the *k*-means clustering algorithm to fulfill its function rapidly, the number of clusters the image should be segmented and starting point of cluster centers are

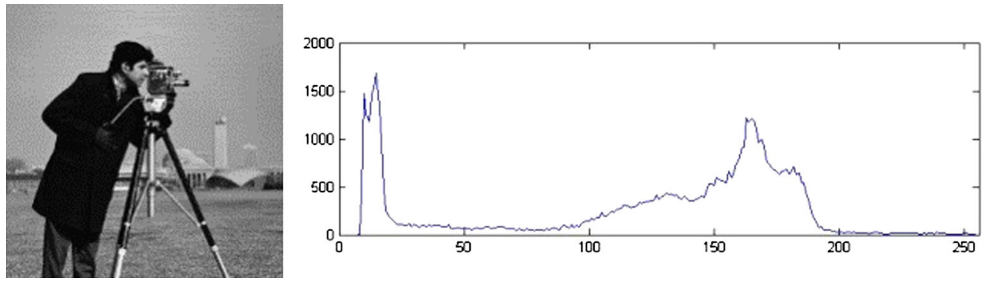


Fig. 1 Cameraman and histogram

Fig. 2 Peak points above average amount of change

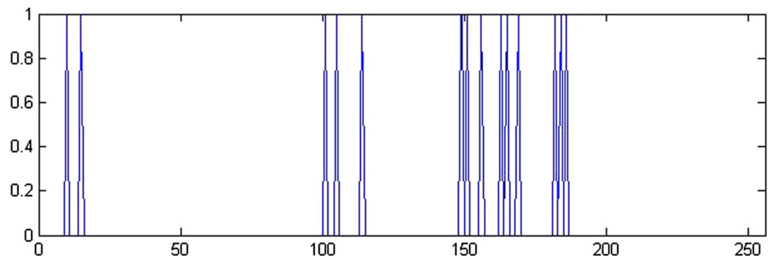
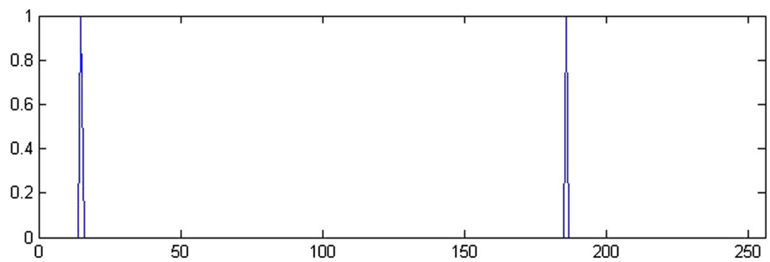


Fig. 3 Cluster starting values determined at the end of algorithm



required apart from image information. In real-life problems, we have no prior knowledge about the number of clusters [5]. In this study, it was aimed to develop a method which predicts the number of clusters and center points by interpreting the histogram of image.

The graphic which shows how many color values there are in a nominal image is called histogram. Brightness and tone information of the image can be obtained through interpretation of histogram of the image [6].

Parameters entered by the user in segmentation algorithms affect the result of segmentation considerably. Benaichouche et al. [7] developed a fuzzy c-means algorithm by using a meta-intuitive optimization technique in which the starting steps are determined automatically. Song et al. [8] presented an automatic training method which uses simplex method in order to choose starting values in page segmentation algorithms. Djerou et al. [9] suggested an automatic multi-level thresholding approach based on a binary particle swarm optimization. In this approach, they suggested gray-level optimization of a function which uses threshold value and determined optimum number of threshold automatically and analyzed

Fig. 4 Segmented cameraman image

optimum threshold values. The algorithm they have developed started with a great number of threshold values and refined threshold values depending on aim function. Cuevas et al. [10] suggested automatic multiple thresholding approach based on differential evolution

Fig. 5 Pseudo-code used for finding cluster counts and centers

```
Img=getImg(),Hist=histogram(Img)(get the histogram of Img)

//determination of peak and pit points in histogram
for i=1:256
        if there is a decrease from increase in values record peak point, else if there is an increase
        from decrease record pit point
end

//scanning of histogram at vertical and exclusion of values smaller than mean value
Calculate mean of distances between each element with the following one at peak and pit
for i=1:length(PeakandPits)
        if the absolute difference between each element with the following one at peak and pit is
        above average, record only peak point at great peaks.
end

//scanning of histogram at horizontal and exclusion of values smaller than mean value
Places=find the places of great peak elements at histogram
for i=1:length(Places)
        Total=absolute distance of each great peak point with the following one*value
end

Mean=total/length of places
for i=1:length(yerler)
        If (absolute distance of each great peak point with the following one*value) is greater than
        the mean, record the great value to the result
end
k= length(result), kStart=result
```

optimization. Here segmentation process was thought to be an optimization problem. In the study, 1d histogram information of the image was filled in by using Gauss function whose parameters are calculated by using differential evolution algorithm. Wen et al. [11] presented automatic statistical concentrated approach in order to obtain 3d cerebrovascular structure from time-of-flight (TOF) magnetic resonance angiography (MRA). They used FFM in order to determine histogram density of brain MR images of structures modeled with Gauss and Rayleigh distribution functions. Güvenç et al. [12] calculated similarity percentage of adjacent pixels by using a mathematical approach based on color similarity measurement. The method they suggested does not require any prior information about possible number of regions in the image. In order to separate color images automatically, Demirci et al. [13] applied Otsu, Kapur and average-based thresholding methods for each channel, and by using threshold values, they restructured three-dimensional color space. For the automatic segmentation of sea lettuce, Qu et al. [14] transformed RGB images into NTSC color space and obtained histogram information and gave two elevation points as cluster center of *k*-means algorithm.

In this study, it was aimed to develop a general approach which would enable automatic segmentation process which is commonly used in image processing without needing any information from user.

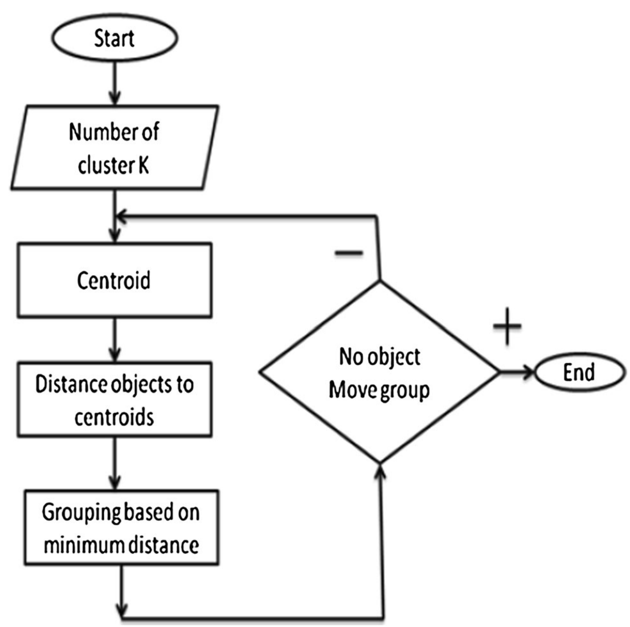


Fig. 6 Simple *k*-means clustering algorithm [15]

2 Determining the number of clusters and cluster centers with histogram

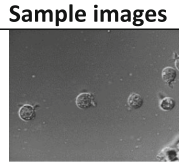
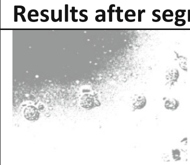
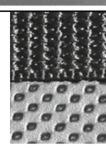
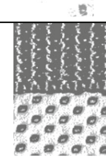
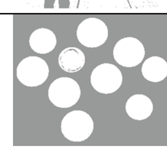
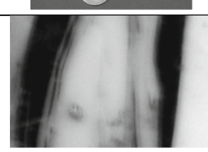
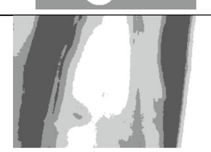
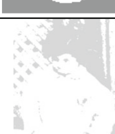
In order to make automatic segmentation process, the number of clusters in which image is to be segmented should be known before conducting segmentation process of images with k-means clustering algorithm. Decision can be taken by looking at histogram of image in order to predict the number of clusters in an image. In Fig. 1, the image of a cameraman and the histogram information of this image are given. By looking at the histogram of image, it can be said that the pixel of the image has color values densified around two basic points.

In order to interpret histogram information, first of all ascending and descending ends on histogram should be determined. In this way, it was enabled to mark peak and pit points on histogram and prepare for the next process.

By calculating the distance between peak and pit, the amount of change was determined and the average amount of change was calculated. At the point of change which is above the average amount of change, peak points which cause this change were determined. In this way, clustering determination can be done even in image that does not have too many tone differences. In Fig. 2, peak points above average amount of change were presented in the cameraman histogram.

The distance of peak points determined with analysis on histogram and values obtained through multiplication of these peak points were calculated. Values, which are above the average, were assigned as cluster starting values in order to be used in clustering process, and the number of them was assigned to k-means clustering algorithm as the number of clusters. In Fig. 3, cluster starting values and number of clusters which were determined in order to assign for k-means algorithm can be seen.

Table 1 Sample images and results after segmentation

Name of samples	Sample images	Results after segmentation
1 AT3_1m4_01.tif		
2 bag.png		
3 cameraman.tif		
4 coins.png		
5 glass.png		
6 moon.tif		
7 mri.tif		
8 pout.tif		

The result obtained by using values derived from algorithm in *k*-means clustering algorithm is presented in Fig. 4.

Pseudo-codes used for finding cluster centers and numbers with the help of histogram are given in Fig. 5.

After the histogram-based algorithm, a flowchart of *k*-means clustering algorithm used for the segmentation process is given in Fig. 6.

3 Simulation results of histogram-based segmentation method

Eight sample images in MATLAB media ('AT3_1m4_01.tif,' 'bag.png,' 'cameraman.tif,' 'coins.png,' 'glass.png,' 'moon.tif,' 'mri.tif' and 'pout.tif') were used in testing the algorithm. Images were first of all sent to the algorithm developed; as a result, number of clusters and starting cluster centers were obtained. Then, these values were sent to *k*-means algorithm, and cluster belonging to pixels was determined. Finally, in distances of 255/number of clusters, pixels were given a color value and results were recorded. Results are given in Table 1.

As a result of tests carried out with sample images, after when peak and pit points are determined in the Step 1 and peak points were determined by horizontal scanning of histogram in the second step, general simplification was done in the histogram. In the Step 3, close values were simplified and distinctive points were marked on histogram. Histogram values obtained after these process steps are given in Table 2; time spent for processes and number of clusters found are given in Table 3.

The results obtained after experiments showed that when the generated algorithm is operated before existing clustering algorithms, algorithm can be made automatic.

Table 2 Histogram of sample images, histogram images after Step 2 and Step 3

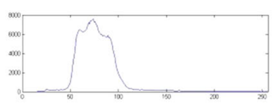
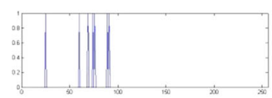
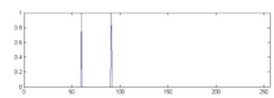
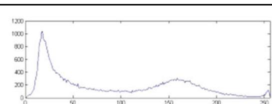
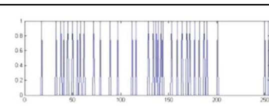
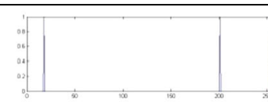
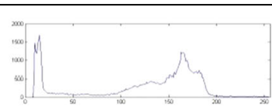
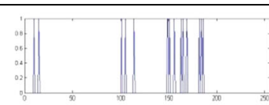
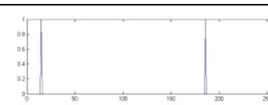
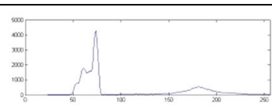
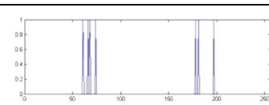
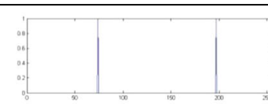
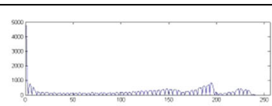
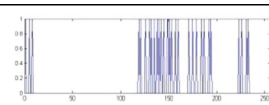
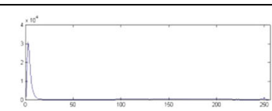
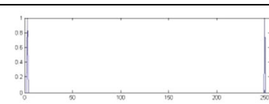
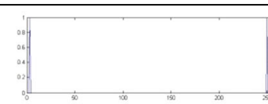
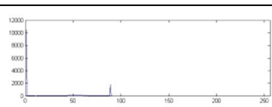
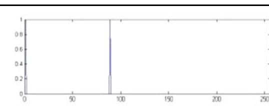
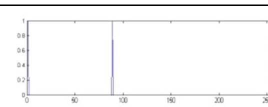
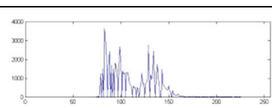
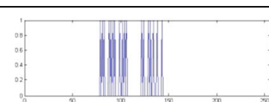
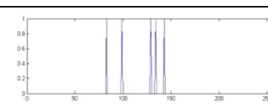
Name of samples	Histogram of images	Scanning of histogram at vertical (Step 2)	Scanning of histogram at horizontal (Step 3)
1 AT3_1m4_01.tif			
2 bag.png			
3 cameraman.tif			
4 coins.png			
5 glass.png			
6 moon.tif			
7 mri.tif			
8 pout.tif			

Table 3 Number of clusters and comparison durations

Number of samples	1	2	3	4	5	6	7	8
Cluster count	2	3	2	2	4	2	2	5
Duration (s)								
Histogram-based algorithm	0.0003	0.0002	0.0001	0.0002	0.0002	0.0003	0.0001	0.0002
Histogram-based algorithm + *k*-means	1.1018	0.1850	0.2367	0.2653	0.1850	0.6890	0.0590	0.2574
k-means	1.0958	0.1840	0.2345	0.262	0.1830	0.6861	0.0596	0.2575

4 Conclusion

In this study, a histogram-based rapid approach was presented in automatic segmentation of images. It was aimed to determine optimum number of clusters and cluster centers through marking peak points by vertical scanning of histogram and exclusion of close values in order not to affect number of clusters and similarly through exclusion of close peaks by horizontal scanning. The number of clusters excluded from histogram and possible center point's information were given to *k*-means clustering algorithm, and segmentation process was done. For the segmentation process success of this suggested method, segmentation process can be done without requiring information such as number and center of clusters. As it is seen in Table 3, histogram-based algorithm provided for preliminary stage of *k*-means clustering process is quite rapid.

References

1. Aydın D (2011) Sürü Zekası YaklaǴımlarının Renkli Görüntü Kesimlemeye Uyarlanması Ve Tanıma Sistemleri Üzerinde GerçekleǴtirimi. Thesis of PhD. Ege Üniversitesi Mühendislik Fakültesi
2. Niknam T, Babak A (2010) An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. Appl Soft Comput 10(1):183–197
3. MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. Proc Symp Math Statist Prob 5:281–297
4. Han J, Kamber M (2006) Data mining concepts and techniques. Morgan Kauffmann, Burlington
5. Kumar V, Chhabra JK, Kumar D (2014) Automatic cluster evolution using gravitational search algorithm and its application on image segmentation. Eng Appl Artif Intell 29:93–103
6. Igarashi M, Mizuno A, Ikebe M (2013) Accuracy improvement of histogram-based image filtering. 20th IEEE international conference on image processing (ICIP), pp 1217, 1221, 15–18. doi:10.1109/ICIP.2013.6738251
7. Benaichouche AN, Oulhadj H, Siarry P (2013) Improved spatial fuzzy c-means clustering for image segmentation using PSO initialization, Mahalanobis distance and post-segmentation correction. Digit Signal Proc 23(5):1390–1400
8. Song M, Kanungo T (2000) Automatic training of page segmentation algorithms: an optimization approach. In: Proceedings of 15th international conference on pattern recognition, vol 4, pp 531, 534. doi:10.1109/ICPR.2000.902974
9. Djerou L, Khelil N, Dehimi HE, Batouche M (2009) Automatic multilevel thresholding using binary particle swarm optimization for image segmentation. International conference of soft computing and pattern recognition, 2009. SOCPAR '09, pp 66, 71, 4–7 Dec. 2009. doi:10.1109/SoCPaR.2009.25
10. Cuevas E, Zaldivar D, Pérez-Cisneros M (2010) A novel multithreshold segmentation approach based on differential evolution optimization. Expert Syst Appl 37(7):5265–5271
11. Wen L, Wang X, Wu Z, Zhou M, Jin Js (2014) A novel statistical cerebrovascular segmentation algorithm with particle swarm optimization. Neurocomputing 148:569–577
12. Güvenç U, Elmas Ç, Demirci r (2008) Renkli Görüntülerin Otomatik Ayrıştırılması. Politeknik Dergisi 11(1):9–12
13. Demirci R, Güvenç U, Kahraman HT (2014) Görüntülerin Renk Uzayı Yardımıyla Ayrıştırılması. İleri Teknoloji Bilimleri Dergisi 3(1):1–8
14. Qu L, Dong X, Guo F (2009) Automatic K-means for color enteromorpha image segmentation. Intell Inf Technol Appl 3:224–227
15. Soumya DS, Arya V (2013) Chromosome segmentation using K-means clustering. Int J Sci Eng Res 4(9):937–940