ISSN : 0974 - 7435 Volume 10 Issue 11

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(11), 2014 [5339-5348]

The study of the determination of the optimal number of image segmentation based on the improved FGS method

Zhe Zhang*, Xin Gao, Haizhou Ma

Information Technology Engineering Department, Yellow River Conservancy Technical Institute, Jingming 47500, Henan, (CHINA)

ABSTRACT

The important section of image analysis is the image segmentation, and the first study of image segmentation is the determination of segmentation number. The best segmentation number has comparatively effect on results of image segmentation and image analysis. This article was based on image segmentation and studied calculation model of its best segmentation image. This text summarized that on the basis of image segmentation technique and clustering analysis to talk about principle of GS AND FGS model and application in image segmentation. And this text put forward the method to determine the best image segmentation number based on improved FGS model. Through the contractive analysis of model principal, we can know that improved model has more obvious robustness on the description of image compactness than GS model and FGS model. Through image of Matlab software system and its achievement platform, the better superiority of improved model on determination of best image segmentation number than GS model and FGS model. In addition, this text did cases analysis on the effects of improved model about noise and smoothing system value to get the conclusion that noise has better effect on application of improved model, and smoothing system value has more obvious superiority on improved model than FGS model.

KEYWORDS

Improved FGS model; Matlab software; Image segmentation number; Clustering analysis.

INTRODUCTION

The continuous development of computer information technology promotes the development of the visual image processing. So called image processing is to deal with image information to meet the demand of people's visual and application requirements. The purpose of image processing is to analyze image, but image analysis tends to be localized. So it needs to part the foreground and the background of image, then to do targeted analysis. Therefore, an important technology of image analysis is image segmentation, and image segmentation algorithms are on the basis of determination of number of segmentation. This article is on the optimum algorithm of determining the number of segmentation of the image analysis, so as to lay a theoretical basis for the development of image segmentation technology.

Many people have done their contribution on study of image segmentation. It is these people's efforts, to some extent, to promote the development of the technology of image analysis. Among them: Zhu Junjie (2014) by studying the principle of image segmentation algorithm and experimental contrast found to mark a watershed segmentation method based on edge features. By setting the spectral characteristics of tags to image segmentation, and edge of the segmentation results of high precision. But there are still serious over-segmentation and owe segmentation, and the mean shift method and eCognition is based segmentation method which is based on spectral characteristics, but the edge of the object segmentation accuracy is poorer. Through the above problems, Zhu designed the improved algorithm, and the feasibility of the improved algorithm is verified by experiment. Wang Jicheng etc. (2007) through gray histogram of the image gradient information to join it in in order to put forward the grey - gradient two-dimensional histogram. This method can effectively control the noise interference, and get more accurate clustering number, and fuzzy clustering can be completely unsupervised. Shi Mei (2007) for problems of image segmentation method based on graph theory made the Normalized Cut criterion and rule of Min - Max Cut image segmentation method combined with threshold binarization image segmentation method respectively, and improved the formula to measure the similarity between pixels. And she has achieved certain effectiveness

This paper on the basis of predecessors' research, aiming at the optimal number of segmentation image segmentation technology research to explore image segmentation algorithm of calculating number based on improved FGS mode to provide theoretical reference for image segmentation technology improvement.

OVERVIEW OF IMAGE SEGMENTATION

In the process of image analysis, involved workers often interest in a certain part of the image. And the premise of image processing in the local forecast analysis results is the image segmentation. People call the prospects that they interest in the above image, and the content of the other parts will be known as the background. Only the background and foreground apart can people more effectively analyze the image. So the process of segmentation is called as image segmentation.

If set R represents whole image area, we can treat image segmentation as the process to divide R into sub areas $R_1, R_2, \cdots, R_n$ that meet the following five conditions:

Condition 1. $\bigcup_{i=1}^{n} R_i = R$

Condition 2. $R_i \cap R_j = \phi; i \neq j$

Condition 3. $P(R_i) = \text{TRUE}; i = 1,2,\cdots,n$

Condition 4. $P(R_i \cup R_j) = \text{FALSE}; i \neq j$

Condition 5. R_i is the connected area

PS: $P(R_i)$ is the logical predicate defined in set R_i.

ϕ represents the empty set.

Scholars at home and abroad did a lot of research on image segmentation, and we can know from those research that the main method of image segmentation can be divided into the method based on threshold value, based on the edge method, the method based on region and the method of combining the theory of specific four classes. Image segmentation technology's development has close relation to mathematics, biology, physics, computer science and many other fields. Due to image segmentation has no general theory so far, people combined with some new theoretical tools to put forward a lot of new image segmentation algorithms, and proposed the improved method to the original segmentation algorithm, such as combining the theory of fuzzy genetic algorithm, mathematical morphology image segmentation algorithm, combined with the fractal geometry, dynamic programming, graph theory tools such as segmentation algorithm, etc.

Zhou Xiancheng (2007) pointed out that the domestic and foreign scholars carried out extensive and in-depth research for image segmentation, and various segmentation method is proposed. But so far there is no complete theory system. Now proposed segmentation methods mostly are focus on the specific problem with no universal applicability. Thus in the future to seek strong adaptability and fast and efficient method will be the research emphasis. In general, image segmentation method is developing towards the direction more quickly, more accurate. Through combining with all sorts of new theories and new technologies it will continue to make breakthroughs and innovations.

Nowadays, the study content of image segmentation technique mainly includes the following two kinds:

(1) To process reasonable image segmentation under the condition of certain segmentation number.

(2) To ensure the reasonable image segmentation number.

The content of this article's study is the best determination method of image segmentation number.

Clustering analysis is the fundamental method to determine the number of image segmentation. This method is derived from mathematics, computer science, statistics, biology and economics, etc. Due to the method showing the advantage in the process of mass data processing, it has been widely used and development. But the method of clustering analysis has shown below yes two significant limitations:

(1) Existing methods can classify only good separating degree data. Now almost all of the algorithms are clustering to class and class non-overlapping data, but if the class and between classes of mutual penetration, then the result of classification will be not very likely expectations.

(2) All clustering method analysis only can analyze the simple one-to-one relationship, ignoring the nonlinear characteristics of clustering.

Common clustering algorithms are mainly divided into the type clustering method, the hierarchical clustering method, the clustering method based on density and grid based clustering method and clustering method based on the model. Therein the principle of the clustering algorithm based on model is by choosing appropriate model, first to correctly estimate the model parameters, then using the rule of the Bayes classifying data. It is a clustering method based on the theory of probability and statistics. Common models of clustering are finite mixture model and Gaussian mixture model, etc.

Tibshirani R, etc. put forward GS model in 2000 to estimate the optimal clustering number. Chen rong, Huang (2005) used the GS method to set image segmentation Gap statistic model which contains regular and singular part. Li Na (2006) changed the Gap statistic, and reduced the complexity of the GS method.

Christopher S.[2] applied the ideas of the Gap statistic to FCM method, and put forward the FGS method. Chinatsu A.[3] improved the model of FGS. FGS method is an extension of the GS method in

fuzzy set theory. If the sample points in the model are the grade of 0 or 1 belonging to one kind, then FGS model is GS model. But when dealing with some vague data clustering, FGS method can better describe the characteristics of the data clustering to give more reasonable optimal clustering number, and image segmentation can be seen as such a cluster, some sample points in the image to the class of membership degree is not so sure, but belonging to different classes in different degree

PRINCIPLE OF GS AND FGS MODEL AND APPLICATION IN IMAGE SEGMENTATION

Principle of GS model and its application in image segmentation

GS represents clustering analysis method. Assumed that $x_{ij}, i = 1,2,\cdots,n;\ j = 1,2,\cdots,p$ is that n independent observing data with p indicators, then making $d_{ll'}$ be the distance square on the space of observing data. If the observing data has been clustered k sorts: $C_1, C_2, \cdots, C_k$. Therein, C_r is the r-th kind, then $n_r = |C_r|$. If making the expression function of D_r and W_k as (1), then defining Gap statistic $Gap_n(k)$ as (2):

$$\begin{cases} \mathbf{D_r} = \sum\limits_{\mathbf{x_l}, \mathbf{x_{l'}} \in \mathbf{d_{ll'}}} \left\| \mathbf{x_l} - \mathbf{x_{l'}} \right\|^2 = \sum\limits_{\mathbf{x_l}, \mathbf{x_{l'}} \in \mathbf{d_{ll'}}} \mathbf{d_{ll'}} \\ \mathbf{W_k} = \sum\limits_{\mathbf{r=1}}^{\mathbf{k}} \left(\dfrac{\mathbf{1}}{\mathbf{2n_r}} \sum\limits_{\mathbf{x_l}, \mathbf{x_{l'}} \in \mathbf{d_{ll'}}} \left\| \mathbf{x_l} - \mathbf{x_{l'}} \right\|^2 \right) = \sum\limits_{\mathbf{r=1}}^{\mathbf{k}} \left(\dfrac{\mathbf{1}}{\mathbf{2n_r}} \mathbf{D_r} \right) \end{cases} \quad \textbf{(1)}$$

$$\mathbf{Gap_n(k) = E_n^* \{ log(W_k) \} - log(W_k)} \quad \textbf{(2)}$$

E_n^* of (2) means expectation of some reference distribution, the best clustering number k appeals when $Gap_n(k)$ picks up the maximum. Reference distribution is the key section of GS method, and uniform distribution can be used very good applied to one dimensional case, for many GS method for the data.

Using singular value (SVD) method to gain main component part of the source data, and then from a set of uniform distribution on the principal component data generating the reference characteristics, the specific algorithm implementation as shown below:

1) To cluster sample X to get k classifications. Calculating every classification's W_k.

2) Using simple uniform distribution to generate B reference data sets, and clustering every data set and calculating its W_{kb}^*, therein $b = 1,2,\cdots,B$. Then we have relationship like (3):

$$\mathbf{E_n} \left[\mathbf{log} \left(\mathbf{W_k^*} \right) \right] = \frac{\mathbf{1}}{\mathbf{B}} \sum_{\mathbf{b=1}}^{\mathbf{B}} \mathbf{log} \left(\mathbf{W_{kb}^*} \right) \quad \textbf{(3)}$$

Using calculation method of (2) to get $Gap_n(k)$.

If it needs image $f(u), u \in S$, 's some at most countable segmentation $A_1, A_1, \cdots \left(S = \sum_{i=1}^{\infty} A_i \right)$, then it needs to definite $\alpha > 0, \beta > 0 (\alpha + \beta = 1)$ the satisfied function (4):

$$\max \quad O(A_i; i \in N) = \alpha G(A_i; i \in N) - \beta H_{n-1}\left(\bigcup_{i \in N} \partial(A_i) \right) \tag{4}$$

$\alpha G(A_i; i \in N) = totalGap\{A_i; i = 1,2,\cdots,\}$ of (4)is the total clearance of function $f(u), u \in S$ about at most segmentation $A_1, A_2, \cdots$. $H_{n-1}\left(\bigcup_{i \in N} \partial(A_i) \right)$ is set $\bigcup_{i \in N} \partial(A_i)$'s Hausdorff measure in $n-1$ dimension. zhang Zhengjun (2005) proposed the Gap statistic model of entropy, (2007) put forward estimates on the number of parts image segmentation of standardization level of Gap statistics, (2009)put forward again the GS model for image segmentation that contained location information and gray information.

Principle of FGS model and its application in image segmentation

FGS is fuzzy clustering analysis method. Defining $d_{ij} = \left\| x_i - x_j \right\|$, written as $W_k = \sum_{i=1}^{n} \sum_{j=1}^{k} \mu_{ij}^m d_{ij}^2$. Therein, k is classification number, W_k reflects compactness of fuzzy clustering. So the definition of $FGap(k)$ is as (5):

$$FGap(k) = E_n\left[\log\left(W_k^* \right) \right] - \log(W_k) \tag{5}$$

$E_n\left[\log\left(W_k^* \right) \right]$ of (5)is expectation of some reference distribution. To calculating the expectation, it need to process B times Monte Carlo simulation, simulating the reference distribution and respectively clustering calculating $W_{kb}^*, (b = 1,2,\cdots,B)$. So the calculation of expectation can get by using formula (3).

FGS method needs B times simulation and fuzzy clustering for the distribution of the reference, so for each k, when calculate the $FGap(k)$ value of the $B+1$ times of fuzzy clustering. The site is home to the disadvantages of fuzzy clustering, according to the ideas of the Gap statistic method, the reference distribution part should be about the class number k is close to the linear function as far as possible. Therefore, in the case of one dimension using uniform distribution as the reference distribution, U_X is set to sample points X corresponding uniform distribution. In order to ensure the symmetry of the model, based on the fuzzy clustering method is as shown in type (6) the iterative formula, defining $\forall x_i^* \in U_X$ its μ_{ij}^* membership for the j class As shown in type (7):

$$v_j^{(b+1)} = \frac{1}{\sum_{k=1}^{n} \left(\mu_{kj}^{(b+1)} \right)^m} \left[\sum_{k=1}^{n} \left(\mu_{kj}^{(b+1)} \right)^m \cdot x_i \right]; i = 1,2,\cdots,c \tag{6}$$

$$\mu_{ij}^* = \frac{1}{\sum_{p=1}^{k} \left(d_{ij}^* / d_{ip}^* \right)^{\frac{1}{n-1}}} \tag{7}$$

For gray image segmentation, if only considering the image gray level information, the gray image segmentation number will be reduced to image grey value of classification number. Using the FGS model to determine the optimal segmentation of image segmentation for the implementation of the steps as follows:

STEP1. Selecting the range of clustering number, $C_{\min} \le k \le C_{\max}$.

STEP2. For $k = C_{\min} \sim C_{\max}$, respectively processing FCM clustering to image grey degree sample X.

STP3. Calculating W_k of every sort of k.

STEP4. According to sample X to generate uniform distribution U_X. For every kind of k, calculating W_k^* with simplified FGS method.

STEP5. Calculating value of $FGap(k)$, then the best clustering number is the minimum k when $FGap(k) > FGap(k+1)$.

IMPROVED FGS MODEL AND IMAGE SEGMENTATION APPLICATION

Model foundation

This article put forward FGS improved model on the basis of study of compactness of fuzzy clustering. Making W_k's expression as (8), calculation of $MFGap(k)$ is as (9):

$$W_k = \max_{j=1}^{k} \sum_{i=1}^{n} \mu_{ij}^m \left\| x_i - x_j \right\|^2 \tag{8}$$

$$MFGap(k) = E_n\left[\log\left(W_k^*\right)\right] - \log\left(W_k\right) \tag{9}$$

For FCM clustering, if $\mu_{ij*} = \max_{j=1}^{k} \mu_{ij}$, and $x_i \in C_{j^*}$. Then W_k has calculation like (10):

$$\mathbf{W_k = \max_{j=1}^{k} \sum_{i=1}^{n} \mu_{ij}^m \left\| x_i - x_j \right\|^2 = \sum_{j=1}^{k} \sum_{x_i \in C_{j^*}} \mu_{ij}^m \left\| x_i - x_j \right\|^2 = \sum_{j=1}^{k} \sum_{x_i \in C_{j^*}} \mu_{ij}^m d_{ij}} \tag{10}$$

Improved W_k, compared with W_k of the GS method, just added a coefficient in front of the distance, so it still can response the firmness of data clustering, and compared with the FGS it only took the largest of the membership degree of fuzzy distance and therefore characterized by more robust clustering firmness. For each x_i if you have $\max_{j=1}^{k} \mu_{ij} = 1$, W_k is calculated as shown in type (11):

$$\mathbf{W_k = \max_{j=1}^{k} \sum_{i=1}^{n} \mu_{ij}^m \left\| x_i - v_j \right\|^2 = \sum_{j=1}^{k} \sum_{x_i \in C_{j^*}} \left\| x_i - v_j \right\|^2 = \sum_{j=1}^{k} \sum_{x_i \in C_{j^*}} d_{ij}} \tag{11}$$

The application of improved FGS model in determination of the optimal number of image segmentation

Combining of clustering analysis and image segmentation translates image segmentation problem into clustering problem, the best segmentation problem into the determination of the optimal clustering number, as shown in Figure 1 improved FGS method best image segmentation framework.

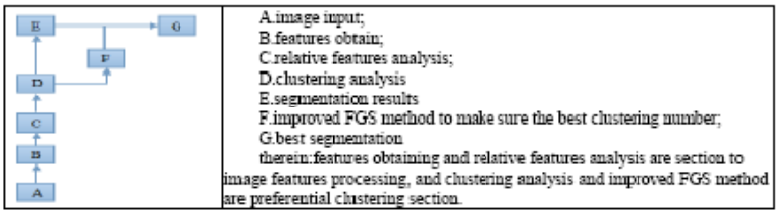
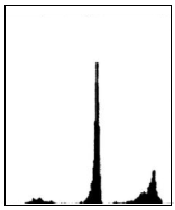


Figure 1 : Frame chart of image best segmentation based on improved FGS model

If using the improved FGS model to segment the inside shot original image of the Matlab software, then the image can be divided into the bull's eye and impact area, the background of the model. Number of the picture of the best segmentation is 3 class as shown in Figure 2 shot straight side of the image Figure 2 2-8 clustering center table.



k \ v	v1	v2	v3	v4	v5	v6	v7	v8
k2	133	315						
k3	64	144	318					
k4	58	92	145	319				
k5	56	83	144	192	223			
k6	53	74	122	145	198	224		
k7	49	62	86	137	147	204	225	
k8	47	61	81	125	145	180	209	226

Figure 2 : shot straight side of the image and 2-8 clustering center tables

$W(k)$ graph and $MFGap(k)$ graph are as Figure 3.

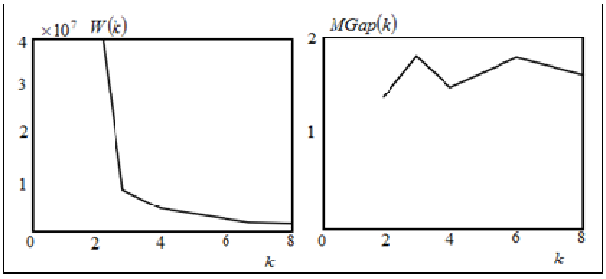


Figure 3 : $W(k)$ graph and $MFGap(k)$ graph

In histogram in Figure 2 shows three distinct peaks, three clustering in the table in the center of the cluster center is 64, 144 and 218 respectively. And the gray level can be found in the $W(k)$ chart the monotonous declining trend curve, the optimal classification number we can easily get from the $MFGap(k)$ curve is 3.

CASES ANALYSIS

As Figure 4, three kinds of curve comparison of GS model, FGS model and improved FGS model in Matlab software system image shot image and cameraman images.

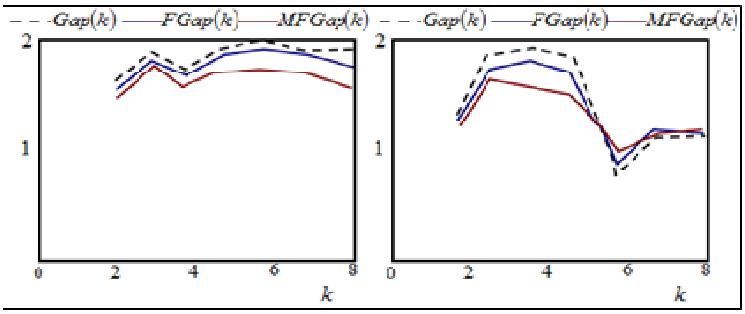


Figure 4 : Comparison of three curves in shot image and cameraman image

Using GS model, FGS model and improved FGS model in Matlab software system image shot image, the cameraman, Lena image and alumgrns images to separately carry on the cluster analysis, the experimental results as shown in TABLE 1.

TABLE 1 : The table of clustering results of three models on fours images

Model class	Shot image	Cameraman image	Lena image	Alumgrns image
GS model	3	4	2	4
FGS model	3	4	4	4
Improved FGS model	3	3	5	4
XB method	3	3	4	4

Ps: Data of table are the final clustering number.

The XB method in TABLE 1 is an index method as well as a common assessment method of effective. Data in the table reflects the XB method and improved FGS model have a better consistency, only in Lena image has a difference. As seen from the histogram, the histogram has five evident in the wave, therefore, it is much better divided into five types.

In order to investigate the robustness of improved FGS model, based on the noise and smooth coefficient situation, this text carries on the analysis of the impact of cluster number. On the Matlab software system in the influence of the noise in the image of Lena image artwork, salt and pepper noise figure and Gaussian noise figure for clustering analysis, the results as shown in TABLE 2. Smoothing coefficient of impact on the number of clustering in the image in this paper, the Matlab software system cameraman, Lena image, testate images and rice image clustering analysis, the results as shown in TABLE 3.

TABLE 2 : The table of results of determination of best segmentation number of three images based on improved FGS model

Clustering number	2	3	4	5	6	7
Lena image	0.982509	1.053373	1.245345	1.392010	1.361380	1.436111
Salt and pepper noise image	1.044251	1.114374	1.265396	1.410433	1.362640	1.485207
Gaussian noise	1.135860	1.162694	1.264685	1.078744	0.948891	1.170677

Ps: Data of the table are $\mathbf{MFGap(k)}$

TABLE 3 : The table of experience results of best segmentation number of four images in different smooth coefficient

Smooth coefficient	Cameraman image		Rice image	
	FGS model	Improved FGS model	FGS model	Improved FGS model

5.0	3	3	3	3
4.5	3	3	3	3
4.0	3	3	3	3
3.5	4	3	3	3
3.0	4	3	4	3
2.5	4	3	4	3
2.0	4	3	4	4
1.5	5	3	4	4

Ps: Data of the table are best segmentation number of images.

From the results of TABLE 2 and TABLE 3, we can get conclusions as following:

1) From TABLE 2, the best segmentation numbers of three images are respectively 5 kinds, 5 kinds and 4 kinds. Original image is similar to salt and pepper noise figure, and is different from the Gaussian noise figure. So the influence of noise on the model is obvious.

2) From TABLE 3, we can know that the image best segmentation number by using improved FGS model in different smooth coefficient value will change.

3) From TABLE 3, different value of smooth coefficient has much bigger effect on FGS model than on improved FGS model, so it proved that the improved FGS model is much more stable and accuracy.

CONCLUSION

First of all, this text on the basis of summarizing the image segmentation method and clustering analysis method in the application of image segmentation number to determine, analyzed the principle of GS model and FGS model and its application in image segmentation, in order to best determine the number of image segmentation model to improve exploration lay the foundation.

Then, on the basis of the GS model and FGS model research, this article put forward the improved model of FGS, and the improved model is more useful than that of GS model to response data of firmness, and more robust than a model of FGS in clustering firmness on the characterization. And this text put forward the research framework in the process of application of the image the best segmentation of improved model, and took shot image in Matlab software system as an example to demonstrate the image segmentation.

Finally, this article using the GS model, FGS model and improved FGS model to process comparison on Matlab software system image shot image and cameraman images, and the curve to show the superiority of the improved model; Using three kinds of model and XB method in Matlab software system image shot images, cameraman, Lena image and alumgrns images separately carried on the cluster analysis, the results showed that the XB method and the improved model has a higher consistency, and the improved model in clustering analysis of Lena image than XB method has more advantages; In TABLE 2 and TABLE 3 in the study of noise and smooth coefficient values of the model, the influence of the noise is relatively obvious, the influence of the model is obtained by smoothing coefficient value of change effects on the improved model is more FGS model has obvious advantages.

ACKNOWLEDGMENT

The key technology research of face recognition based on information campus (2014KXJS010).

REFERENCE

[1] R.Tibshirani, G.Walther, T.Hastie; Estimating the number of cluster in a dataset via the Gap statistic [J]. J.R.Statist.Soc.B., **2(63)**, 411-423 **(2001)**.

[2] S.Christopher, L.H.Siu, G.Michael; A Fuzzy Gap Statistic for Fuzzy C-Means[J]. Artificial Intellengence and Soft Computing, 68-73 **(2007)**.

[3] Chinatsu Arima, Kazumi Hakamada, Masahiro Okamoto; Modified Fuzzy Gap Statistic for Estimating Preferable Number of Cluster in Fuzzy K-Means Clustering [J]. Journal of Bioscience, **105(3)**, 273-281 **(2008)**.

[4] Zhang Zheng-Jun, Zhu Yao-Qin; Estimating the Image Segmentation Number via the Entropy Gap Statistic [J]. World Journal of Modelling and Simulation, **25**, 223-225 **(2005)**.

[5] Zhang Zheng-Jun, Liu Li-Wei, Zhu Yao-Qin; GS Model for the Image Segmentation Number via Weighted the Position Information and the Gray Values [J]. Image and signal processing, **12**, 124-128 **(2009)**.