SIViP (2011) 5:257–267
DOI 10.1007/s11760-010-0159-7

ORIGINAL PAPER

Fast medical image mixture density clustering segmentation using stratification sampling and kernel density estimation

Cong-Hua Xie · Yu-Qing Song · Jian-Mei Chen

Received: 29 August 2009 / Accepted: 3 March 2010 / Published online: 26 March 2010

Abstract The Gaussian mixture models (GMMs) is a flexible and powerful density clustering tool. However, the application of it to medical image segmentation faces some difficulties. First, estimation of the number of components is still an open question. Second, the speed of it for large medical image is slow. Moreover, GMMs has the problem of noise sensitivity. In this paper, the kernel density estimation method is used to estimate the number of components K, and three strategies are proposed to improve the segmentation speed of GMMs. First, a histogram stratification sampling strategy is proposed to reduce the size of the training data. Second, a binning strategy is proposed to search the neighbor points of each center data to compute the approximate density function of the samples. Third, a hill-climbing algorithm with the dynamic step size is designed to find the local maxima of the density function. The kernel density estimation method and sampling technology reduce the effect of noise. Experimental results with the simulated brain images and real CT images show that the proposed algorithm has better performance in generating explainable segmentations with faster speed than the common GMMs algorithm.

C.-H. Xie · Y.-Q. Song · J.-M. Chen
School of Computer Science and Telecommunication,
Jiangsu University, 212013 Zhenjiang, Jiangsu Province,
Peoples' Republic of China

C.-H. Xie (✉)
School of Computer Science and Engineering,
Changshu Institute of Technology,
215500 SuZhou, Jiangsu Province,
Peoples' Republic of China
e-mail: x7c8h5@yahoo.com.cn

Y.-Q. Song
e-mail: yqsong@ujs.edu.cn

J.-M. Chen
e-mail: cjm@ujs.edu.cn

Keywords Expectation maximization · Parzen density estimation · Stratification sampling · Binning data · Hill-climbing algorithm

List of Symbols

X	A medical image
$\{x_1, x_2, \ldots x_N\}$	All pixels of X
N	The number of pixels
N_h	The size of the hth stratum of X
n_h	Size of the hth stratum of SX
W_h	Proportion of the hth stratum of X
w_h	Proportion of the hth stratum of SX
N	The Size of sample SX
K	The number of components of mixture models
α_r	Is the weight of component r
μ_r	Is the mean of component r
$\sum_r$	The covariance of component r
$\varphi_r = \{\alpha_r, \mu_r, \Sigma_r\}$	All parameters of component r
$\theta = (\alpha_1, \alpha_2, \ldots, \alpha_K, \mu_1, \mu_2, \ldots \mu_K, \sum_1, \sum_2, \ldots \sum_K)$	All parameters of mixture model
$f_r(x\|\mu_r, \sum_r)$	The density function of component r
$\log L(x\|\theta)$	Log-likelihood function
SX	The samples extracted from a medical image X
$SY = \{Y_1, \ldots, Y_n\}$	2D data set converted from the 3D data set SX
$G = \{L_{\min}, L_{\min}+1, \ldots, L_{\max}\}$	The gray level set of a medical image X
p	The number of sampling strata
His	The histogram of image X
$SHis$	The smoothed histogram

$\overline{Y}$	The mean of image X
$\overline{y}_h$	The mean of the hth stratum of sample SX
α	The precision
δ_i	The ith step optimum size of hill-climbing procedure
$\tilde{f}(Y_i)$	Approximate density
$\nabla \tilde{f}(Y_i)$	The gradient function
$\nabla^2 \tilde{f}(Y_i)$	Hessian matrix of density function
S	Unit gradient vector
$N_{p\cap g}(r)$	The number of pixels classified by both the proposed method and the ground truth as model r
$N_p(r)$	The number of pixels classified as model r by the proposed method
$N_g(r)$	The number of pixels classified as model r by the ground truth
ξ	Thresholding
σ	Window width
xw	The bin width of x-coordinate
yw	The bin width of y-coordinate
Y	The data in the same bin or neighbor bins of Y_i
s	The size of Y

1 Introduction

Medical imaging is an effective diagnostic tool in today's clinical practice. Medical image segmentation is a critical step toward the content analysis and image understanding, such as the quantification of tissue volumes, the study of anatomical structure and computer-integrated surgery [1]. Due to the presence of noise, intrinsic tissue variation, partial volume effects, unclear tissue boundaries and intensity non-uniformity, medical image segmentation remains a challenging task [2].

There are a lot of methods available for medical image segmentation [3,4], such as specific probability density distribution [5], decision tree [6] and neural networks [7]. In the wide range of segmentation methods, clustering algorithms are termed unsupervised classification methods that organize unlabeled feature vectors into clusters or "natural groups" so that the samples within a cluster are more similar to each other than the samples belonging to different clusters [8]. Most unsupervised clustering techniques [9], including statistical-based clustering [2,5,8], neural network–based clustering [10] and various fuzzy clustering [1,11–13], have been used to accomplish this task. Some of those techniques are reviewed in the following part.

Among the fuzzy clustering methods, fuzzy c-means (FCM) algorithm [11] is the most popular used in image segmentation. Cai [12] proposed a novel fast and robust FCM framework for image segmentation and a fast generalized fuzzy c-means clustering algorithms by incorporating local spatial and gray information together. Yang [13] proposed an improved fuzzy c-means algorithm, incorporating the spatial neighborhood information into the original FCM algorithm by a priori probability. Those fuzzy clustering methods estimate the parameters, which minimize the distance from each voxel to the class centers. They use only the distance objective function without any other information about the intensity distributions. The most important is that they have no ability to estimate the number of components K for medical image segmentation.

In contrast, the GMMs methods use the statistical theory to model each voxel's intensity, which is more reasonable to the real situation. Many researchers [14–16] used the expectation-maximization (EM) algorithm [17] and the frequently used mixture model clustering algorithm to segment medical image. Wells [15] reported an adaptive segmentation method, which applied the EM algorithm on intensity in-homogeneities of different tissues to segment brain medical images. In order to solve the noise sensitivity problem of the GMMs methods, Tang [18] used a multi-resolution GMMs method for image segmentation. Yan [19] used the spatial information as a prior knowledge of the number of components in the application of finite mixtures to image segmentation. Rasoul Khayati [20] proposed a segmentation approach for Magnetic Resonance (MR) images, based on a Bayesian classifier, and utilized the adaptive mixtures method and Markov random field model to obtain and upgrade the class conditional probability density function and the priori probability of each class. Adelino [21] proposed a Dirichlet process mixture model for tissue classification of MR images. The model uses Dirichlet process to overcome the limitations of current parametric finite mixture models. However, the application of GMMs to medical image segmentation faces some difficulties: (1) it is still an open question to estimate the number of components K. Although many model selection criteria [22] are available to estimate it, many of them become invalidated for medical image. The main reason is that the absolute value of log-likelihood function is often far larger than its punishment function, and the image is over-fitting. (2) For the original medical image with large size, the GMMs method has a slow segmentation speed. (3) The GMMs method has the problem of noise sensitivity.

Our motivation lies in the development of fast procedures for the automatic segmentation of medical image, which can determine the number of components K and reduce the effect of noise. On the one hand, a non-parametric density estimation method is used to estimate the number of components as prior knowledge of the GMMs method. We employ the Parzen window approach [23] to estimate the medical image density and propose a hill-climbing algorithm with the dynamic step size to estimate the number of components K. The kernel density estimation is insensitive to noise. On the other hand, we propose three strategies to improve the speed of image segmentation. First, the histogram stratification sam-

pling strategy is proposed to extract the samples. It reduces greatly time and space of estimating the parameters of GMMs and K. Moreover, the sampling strategy reduces the effect of noise. Second, the fast approximate density of image is estimated by replacing the influence functions of all data with local data. We propose the binning strategy to search the local neighbor points to compute the approximate density of the samples. Third, the hill-climbing algorithm with the dynamic step size is proposed to find the local maxima of the density function.

The outline of this paper is as follows. Section 2 introduces the GMMs. Section 3 describes our segmentation method with stratification sampling and kernel density estimation. Experimental results and comparisons are given in Sect. 4. Finally, some conclusions are drawn in Sect. 5.

2 Gaussian mixture models

Suppose a medical image X has N pixel points $\{x_1, x_2, x_3 \ldots x_N\}$ and K components, the finite mixture density of any point x is:

$$f\left(x|\theta\right) = \sum_{r=1}^{K} \alpha_r f_r\left(x|\mu_r, \Sigma_r\right) \left(x \in R^d\right) \tag{1}$$

where $\theta = (\alpha_1, \alpha_2, \ldots, \alpha_K, \mu_1, \mu_2, \ldots \mu_K, \Sigma_1, \Sigma_2, \ldots \Sigma_K)$, α_r is the weight of component r with $\sum_{r=1}^{K} \alpha_r = 1$ and $\alpha_r > 0$. And μ_r and Σ_r are the mean and covariance of component r, respectively. The density function of each component of GMMs is:

$$f_r\left(x|\mu_r, \Sigma_r\right) = \frac{\exp\left[-\frac{1}{2}\left(x-\mu_r\right)^T \Sigma_r^{-1}\left(x-\mu_r\right)\right]}{(2\pi)^{d/2}\left|\Sigma_r\right|^{1/2}} \tag{2}$$

The usual choice for obtaining maximum likelihood estimation of the mixture parameters is the EM algorithm, which finds the local maxima of log-likelihood function:

$$\log L\left(x|\theta\right) = \sum_{i=1}^{N} \log\left(\sum_{r=1}^{K} \alpha_r f_r\left(x_i|\mu_r, \Sigma_r\right)\right) \tag{3}$$

Suppose $\varphi_r = \{\alpha_r, \mu_r, \Sigma_r\}$, the EM algorithm for GMMs is described as below:

Step1. Compute the posterior probability of $p\left(\varphi_r|x_i\right)\left(1 \le i \le N, 1 \le r \le K\right)$:

$$p\left(\varphi_r|x_i\right) = \alpha_r f_r\left(x_i|\mu_r, \Sigma_r\right) \Big/ \sum_{r=1}^{K} \alpha_r f_r\left(x_i|\mu_r, \Sigma_r\right) \tag{4}$$

Step2. Compute the weight, mean and covariance matrix:

$$\alpha_r = \sum_{i=1}^{N} p\left(\varphi_r|x_i\right)/N \tag{5}$$

$$\mu_r = \sum_{i=1}^{N} p\left(\varphi_r|x_i\right)x_i \Big/ \sum_{i=1}^{N} p\left(\varphi_r|x_i\right) \tag{6}$$

$$\Sigma_r = \sum_{i=1}^{N} p\left(\varphi_r|x_i\right)\left(x_i-\mu_i\right)\left(x_i-\mu_i\right)^T \Big/ \sum_{i=1}^{N} p\left(\varphi_r|x_i\right) \tag{7}$$

Step3. Update the posterior probability with formula (5), (6) and (7) and compute the log-likelihood function;

Step4. If the increase in the value of $\log L(x|\theta)$ at the current iteration relative to the value of $\log L(x|\theta)$ at the previous iteration is below a chosen threshold, then the algorithm stops; otherwise, go to Step 2.

3 Our methods

The framework of our method is shown in Fig. 1, in which the samples extracted from a medical image X are labeled as SX. Both X and SX have 3D features of x-coordinate, y-coordinate and gray. First, the histogram stratification sampling algorithm is proposed to extract the samples SX and map SX into SY that has 2D features of x-coordinate and y-coordinate by the multidimensional scaling method. Second, the approximate density of SY is estimated by the binning strategy. Third, the number of components K is estimated by the hill-climbing algorithm with the dynamic step size. Finally, image X is segmented by the K models of SY.

3.1 Histogram stratification sampling

The computation of Parzen density estimation is directly related to the number of training samples. Therefore, the density estimation of medical image with large size is too computationally intensive. A histogram stratification sampling method is proposed to reduce the size of training data. Some notations of our sampling algorithm are listed in Table 1.

3.1.1 Histogram-based strata

The strata should be internally as homogeneous as possible, because the smaller strata variance, the smaller is the variance of the sample estimation [24]. Histogram threshold based image segmentation methods show that the histogram is a natural standard to form image stratum. Suppose the gray level set of image X be $G = \{L_{\min}, L_{\min}+1, \ldots, L_{\max}\}$ ($L_{\min}$ = the minimum gray level, $L_{\max}$ = the maximum gray level) and gray (x_i) be the gray value of point x_i, the

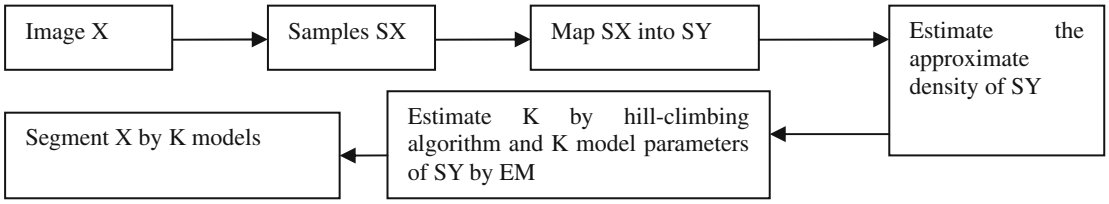

Fig. 1 The framework of our segmentation method

Table 1 Notations of our sampling algorithm

Notations	N	n	N_h	n_h	$W_h = N_h/N$	$w_h = n_h/N_h$
Explanation	Size of image X	Size of sample SX	Size of the hth stratum of X	Size of the hth stratum of SX	Proportion of the hth stratum of X	Proportion of the hth stratum of SX

histogram of image X is:

$$\text{His (g)}=\sum_{i=1}^{N}\delta(gray(x_i)-g)(g\in G) \tag{8}$$

here $\delta(0)=1$ and $\delta(g\neq 0)=0$. For $His(g)$ is not smooth, we replace it with its smoothed histogram $SHis(g)$. Suppose the smoothed histogram $SHis(g)$ has the local minima at the gray levels of $T_1, T_2, \ldots T_{p-1}$. The gray level set G can be divided into p subsets: $G_1 = [L_{min}, \ldots, T_1], G_2 = [T_1, \ldots, T_2], \ldots, .G_p = [T_{p-1}, \ldots, L_{max}]$.Those pixels whose gray values belong to one of those p subsets form one stratum. Image X is decomposed into p strata.

3.1.2 Estimation of the sizes of each strata and total sample

The samples SX are extracted by independently random sampling from each stratum of image X. The sample size of each stratum of SX is proportionate to the population size of each stratum of X. The size of the hth stratum of SX is:

$$n_h=\frac{n}{N}N_h \quad (1\le h\le p) \tag{9}$$

We label the gray of the ith pixel of the hth stratum as $y_{h,i}$, the mean of image X as $\overline{Y}$ and the mean of sample SX as $\overline{y}$. The mean of the hth stratum of image X is:

$$\overline{Y}_h=\sum_{i=1}^{N_h}y_{h,i}/N_h \tag{10}$$

and the mean of the hth stratum of sample SX is:

$$\overline{y}_h=\sum_{i=1}^{n_h}y_{h,i}/n_h. \tag{11}$$

The variance of the hth stratum of image X is:

$$S_h^2=\sum_{i=1}^{N_h}(y_{h,i}-\overline{Y}_h)^2/(N_h-1). \tag{12}$$

An unbiased estimate of the variance of $\overline{y}$ is $\text{var}(\overline{y}) = S^2(\overline{y}) = \sum_{h=1}^{p}\frac{W_h^2S_h^2}{n_h} - \sum_{h=1}^{p}\frac{W_hS_h^2}{N}$. And because $n_h = n.w_h(h=1,2,\ldots,p)$, we have:

$$\text{var}(\overline{y})=\frac{1}{n}\sum_{h=1}^{p}\frac{W_h^2S_h^2}{w_h}-\sum_{h=1}^{p}\frac{W_hS_h^2}{N} \tag{13}$$

If the margin of error is less than d, the precision requires $P(|\overline{y}-\overline{Y}|\le d)=1-\alpha$. We have

$$P\left[\left|\frac{\overline{y}-\overline{Y}}{\sqrt{\text{var}(\overline{y})}}\ge\frac{d}{\sqrt{\text{var}(\overline{y})}}\right|\right]\ge 1-\alpha \tag{14}$$

And note $\frac{d}{\sqrt{\text{var}(\overline{y})}}=\mu_{\alpha/2}$, that is, $\text{var}(\overline{y})=d^2/(\mu_{\alpha/2})^2$, where $\mu_{\alpha/2}$ is the right $\alpha/2$ quantile of the standard normal distribution. For the proportionate stratification, we have $W_h = w_h$. According to formula (13), the size of sample SX is:

$$n=\frac{\sum_{h=1}^{p}W_hS_h^2}{\frac{d^2}{(\mu_{\alpha/2})^2}+\frac{1}{N}\sum_{h=1}^{p}W_hS_h^2} \tag{15}$$

Our stratification sampling algorithm is listed as below.

Algorithm 1: Histogram-based stratification sampling

Input: Image X, the margin of error d
Output: Sample SX
(1) Compute His (g) of X, $g\in G$;
(2) Compute the smooth histogram SHis(g) of X, $g\in G$;
(3) p = the count of local minima of SHis;
(4) Image X is divide into p strata;
(5) Compute the total size of sample n by formula (15);
(6) For(h = 1; h <= p; h + +)
(7) Random sample n_h data from the hth stratum and label as SX_h;
(8) $SX = SX \cup SX_h$.

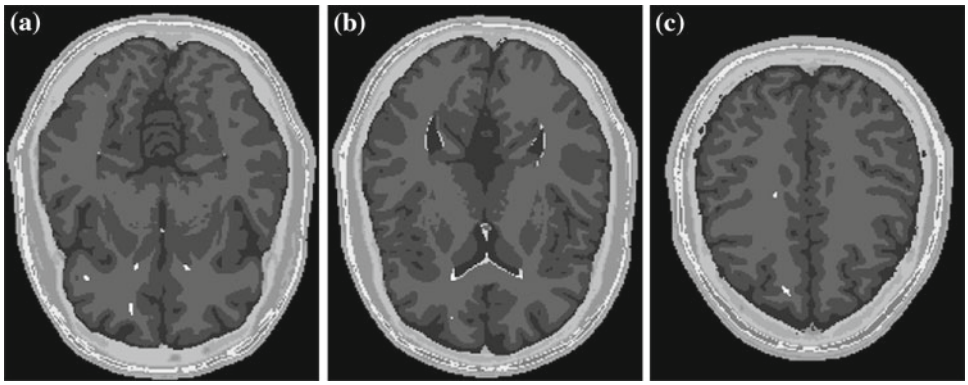

Fig. 2 The BrainWeb images for different axial slices. **a** Slice no. 66; **b** Slice no. 86; **c** Slice no. 116

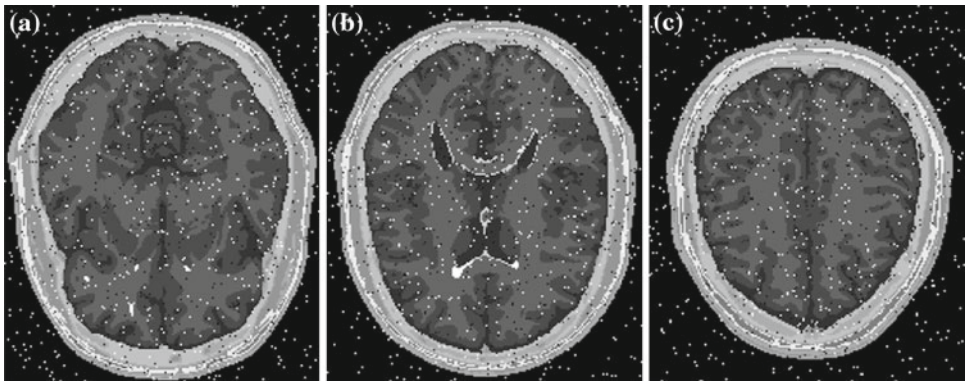

Fig. 3 The BrainWeb images for different axial slices with 3% noise. **a** Slice no. 66; **b** Slice no. 86; **c** Slice no. 116

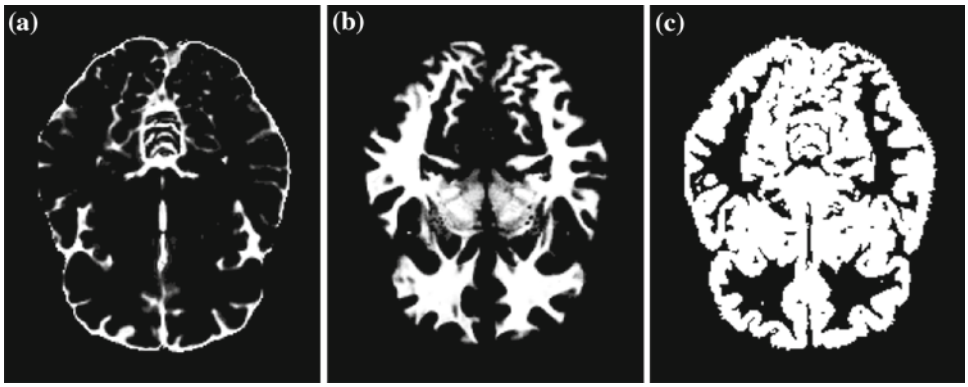

Fig. 4 The ground truth of CSF, WHT, GRY of Fig. 2a. **a** CSF; **b** WM; **c** GM

3.2 Approximate density estimation with binning strategy

Density of medical image X is the construction of an estimate, based on the observed data SX, of an unobservable underlying probability density function. Each sample point is a 3D data with the coordinates and gray. Because the gray of the unobservable data is unknown, we need convert the 3D data set SX into 2D data set $SY = \{Y_1, \ldots, Y_n\}$ that can be displayed in a scatter plot. In the mapped 2D space, we can estimate the density of any data. Dimensionality reduction methods can do it. The aim of dimensionality reduction is to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map [25]. Various techniques that differ in the type of structure for this problem have been proposed. Classical multidimensional scaling (MDS) is used to map SX into SY in this paper.

The Parzen density function of each data is defined as the sum of the influence functions of all data. In $SY = \{Y_1, \ldots, Y_n\} \in R^2$, the density function of Y_i $(1 \leq i \leq n)$ is:

$$f(Y_i) = \frac{1}{n\sigma} \sum_{j=1}^{n} K\left(\frac{Y_j - Y_i}{\sigma}\right), \tag{16}$$

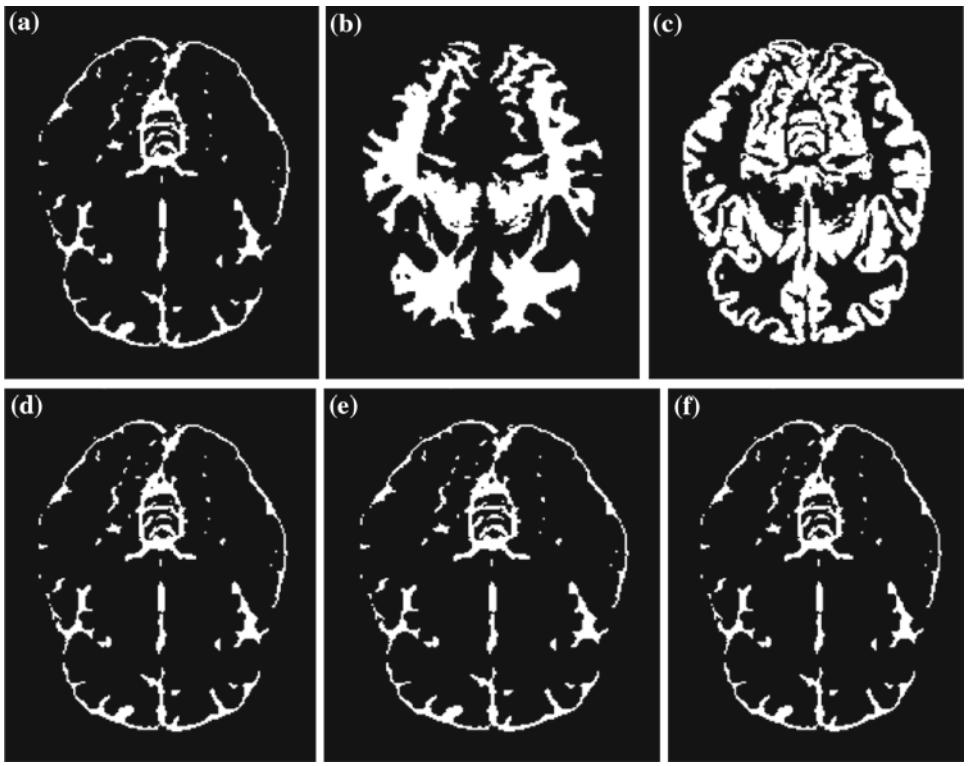

Fig. 5 The segmentation of CSF, WHT, GRY of Fig. 2a; **a** CSF by our method; **b** WM by our method; **c** GM by our method; **d** CSF by C-FCM; **e** CSF by F-GMMs; **f** CSF by Y-GMMs

Table 2 The average DSM index and time of the 10 BrainWeb images

	Time(s)	DSM for images			DSM for images with 3% noise		
		CSF	WM	GM	CSF	WM	GM
C-FCM	1.87	0.93	0.88	0.85	0.89	0.85	0.84
F-GMMs	3.65	0.91	0.85	0.83	0.85	0.82	0.80
Y-GMMs	3.43	0.93	0.92	0.90	0.84	0.81	0.81
Our method	0.38	0.94	0.92	0.91	0.93	0.91	0.90

where $K(.)$ is a kernel function and σ is called as window width. Although formula (16) can compute the density function of SY, its speed is slow. A faster density estimate method is available by only considering the influence functions of local data instead of all data. We propose a binning strategy to search the neighbors of each data.

Let the maximum and minimum of new x-coordinate and y-coordinate of SY be $x_{\max}, x_{\min}, y_{\max}$ and $y_{\min}$, respectively. All of the n samples are partitioned into the $xn \times yn$ bins, where xn and yn are the number of bins in x-coordinate and y-coordinate, respectively. Suppose $xn = yn = \sqrt{n}$, the bin width of x-coordinate is:

$$xw = (x_{\max} - x_{\min})/\sqrt{n} \tag{17}$$

The bin width of y-coordinate is:

$$yw = (y_{\max} - y_{\min})/\sqrt{n} \tag{18}$$

Then, we can compute the location of the bin for any data $Y_i(x, y)$ by:

$$bi = \lfloor \left(\frac{x - x_{\min}}{xw} \right) \rfloor + 1 \tag{19}$$

$$bj = \lfloor \left(\frac{y - y_{\min}}{yw} \right) \rfloor + 1 \tag{20}$$

Suppose a set Y comprises the data in the same bin or neighbor bins of Y_i.The approximate density of Y_i is defined as the sum of influence functions of each data in Y. The approximate density function of any data with Gaussian

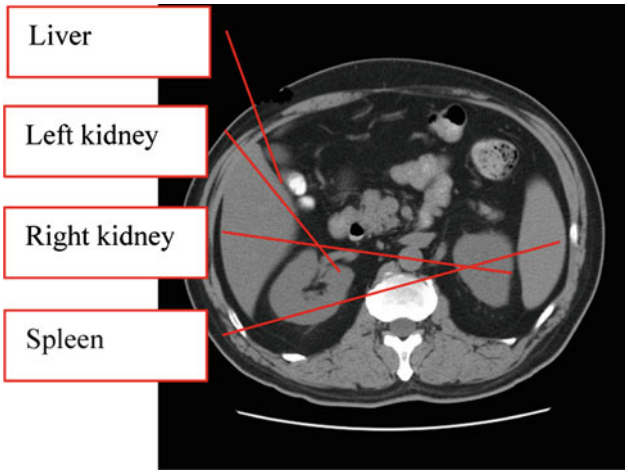

Fig. 6 Original CT abdomen image X

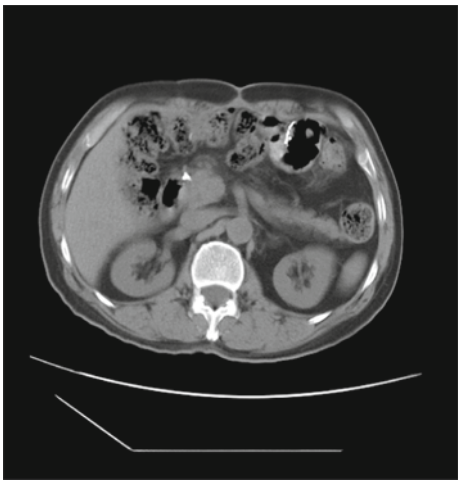

Fig. 7 Another CT abdomen image

kernel function is:

$$\tilde{f}(Y_i) = \frac{1}{(\sqrt{2\pi}\sigma)^2} \sum_{Y_j \in Y} e^{\frac{-|Y_j - Y_i|^2}{2\sigma^2}} \tag{21}$$

The smooth parameter σ shows the influence degree of a data to the others. We use the average distance of the neighbor bins as the value of σ. If the size of Y is s, the smooth parameter is:

$$\sigma = \sum_{Y_j \in Y} \sqrt{(Y_j - Y_i)(Y_j - Y_i)^T}/s \tag{22}$$

In summary, the fast approximate density estimation algorithm is listed as following:

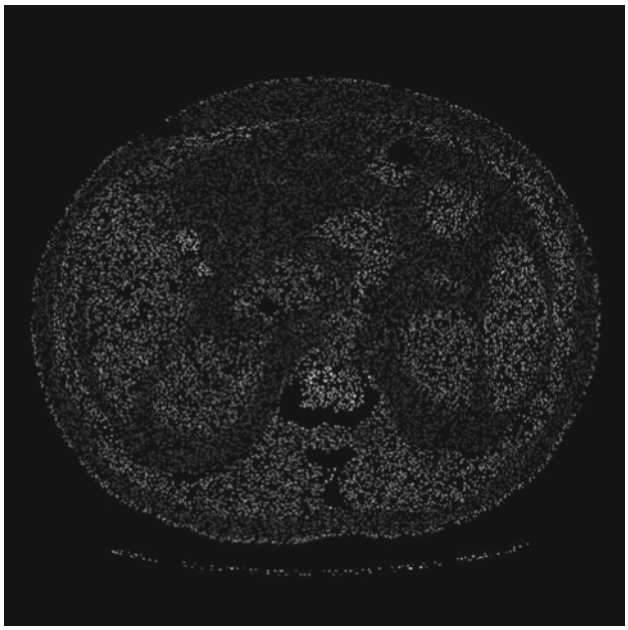

Fig. 8 Sample image SX

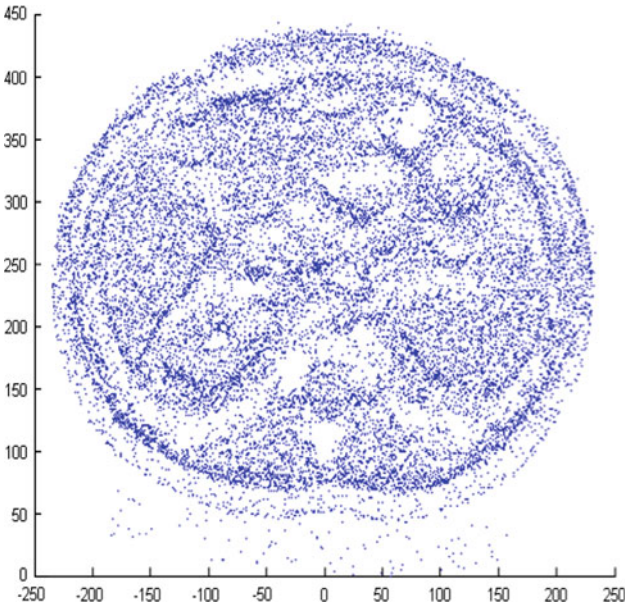

Fig. 9 Mapped 2-dimensional image by MDS

Algorithm 2: Approximate Kernel density estimation

Input: $SX = \{X_1, \ldots, X_n\} \in R^3$
Output: Approximate Kernel Density $\tilde{f}$
(1) Map $SX \in R^3$ into $SY = \{Y_1, \ldots, Y_n\} \in R^2$ by MDS;
(2) SY is loaded into the bins according to formula (17) and (18);
(3) For each data Y_i in SY;
(4) Compute its bin location by (19) and (20) and find its neighbors;
(5) Compute the smooth parameter σ;
(6) Compute the approximate density by (21).

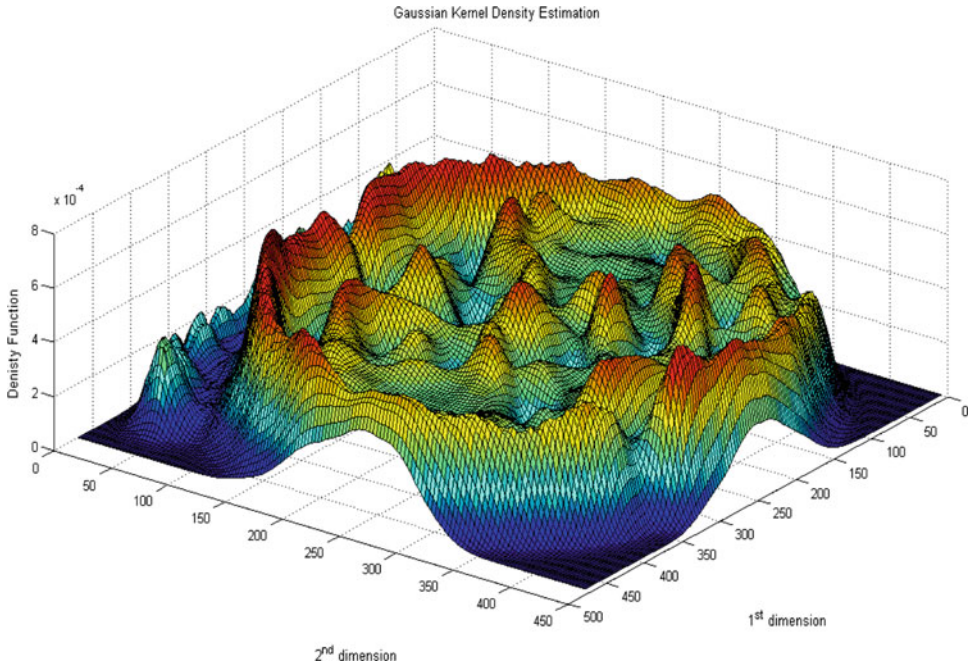

Fig. 10 Approximate kernel density function of Fig. 8 ($n = 19{,}624$)

3.3 Determine K by hill climbing with the dynamic step size

The approximate density of SY can express the clusters or components. Data points are assigned to the local maxima by hill climbing. Those points that are assigned to the same local maximum are put into a single cluster. Each "hill" of the density is a component of image. The number of local maxima of density function is the number of components K. Hill-climbing algorithm can find those local maxima with the guide of gradient function. The gradient function at Y_i is:

$$\nabla \tilde{f}(Y_i) = \sum_{Y_j \in Y} (Y_i - Y_j)\tilde{f}(Y_i) \tag{23}$$

The hill climbing converges toward the local maximum Y*. All points on the path from Y_i to Y* are treated as outliers or noise with $\tilde{f}(Y^*) < \xi$. Otherwise, those points are assigned to Y*. In order to find those local maxima, we propose the hill climbing with the dynamic step size. Denote Hessian matrix of density function as $\nabla^2 \tilde{f}$, the ith step size as δ_i and unit gradient vector $S_i = \frac{\nabla \tilde{f}(Y_i)}{\|\nabla \tilde{f}(Y_i)\|}$ in the hill-climbing procedure.

Lemma 1 *The ith step optimum size of hill-climbing procedure is* $\delta_i = \frac{-\nabla^T \tilde{f}(Y_i) S_i}{S_i^T \nabla^2 \tilde{f}(Y_i) S_i}$.

Proof Suppose the iterative formula in hill-climbing procedure is $Y_{i+1} = Y_i + \delta_i \cdot S_i$. We have the density of Y_{i+1} is $\tilde{f}(Y_i + \delta_i S_i)$. $\tilde{f}(Y_{i+1})$ is unfolded at Y_i by Taylor formula, then $\tilde{f}(Y_{i+1}) = \tilde{f}(Y_i) + \nabla^T f(Y_i) \cdot \delta_i S_i + \frac{1}{2}(\delta_i S_i)^T \nabla^2 f(Y_i)(\delta_i S_i)$. And the derivatives of density function $\frac{\partial \tilde{f}}{\partial \delta_i} = \nabla^T \tilde{f}(Y_i) S_i + \delta_i S_i^T \nabla^2 \tilde{f}(Y_i) S_i$. Because $\tilde{f}(Y_i)$ has the extremum at Y_{i+1}, $\frac{\partial \tilde{f}}{\partial \delta}\big|_{\delta=\delta_i} = \nabla^T \tilde{f}(Y_i) S_i + \delta_i S_i^T \nabla^2 \tilde{f}(Y_i) S_i = 0$. Therefore, we have $\delta_i = \frac{-\nabla^T \tilde{f}(Y_i) S_i}{S_i^T \nabla^2 \tilde{f}(Y_i) S_i}$. □

Table 3 The number of clusters K with different size of sample $n(\xi = 2 \times 10^{-4})$

n	7000	9000	11000	13000	15000	17000	19000	21000
K	16	17	17	17	18	18	18	18

According to Lemma 1, our hill-climbing algorithm is listed as below.

Algorithm 3: Hill-climbing algorithm with the dynamic step size

Input: Approximate density $\tilde{f}$, thresholding ξ
Output: The number of components K
(1) K=0;
(2) while ($SY \neq \phi$)
(3) { Y_i=Select any data from SY;
(4) $SY = SY - Y_i$;
(5) Compute $\nabla \tilde{f}(Y_i)$ and δ_i;
(6) while ($\|\nabla \tilde{f}(Y_i)\| > \xi$ and $\tilde{f}(Y_{i+1}) > \tilde{f}(Y_i)$)
(7) { $Y_{i+1} = Y_i + \delta_i \cdot S_i$;
(8) $Y_i = Y_{i+1}$;
(9) $SY = SY - Y_i$;}
(10) $Y^* = Y_i$;
(11) if ($\tilde{f}(Y^*) \geq \xi$) and Y^* is not visited
(12) K=K+1; }

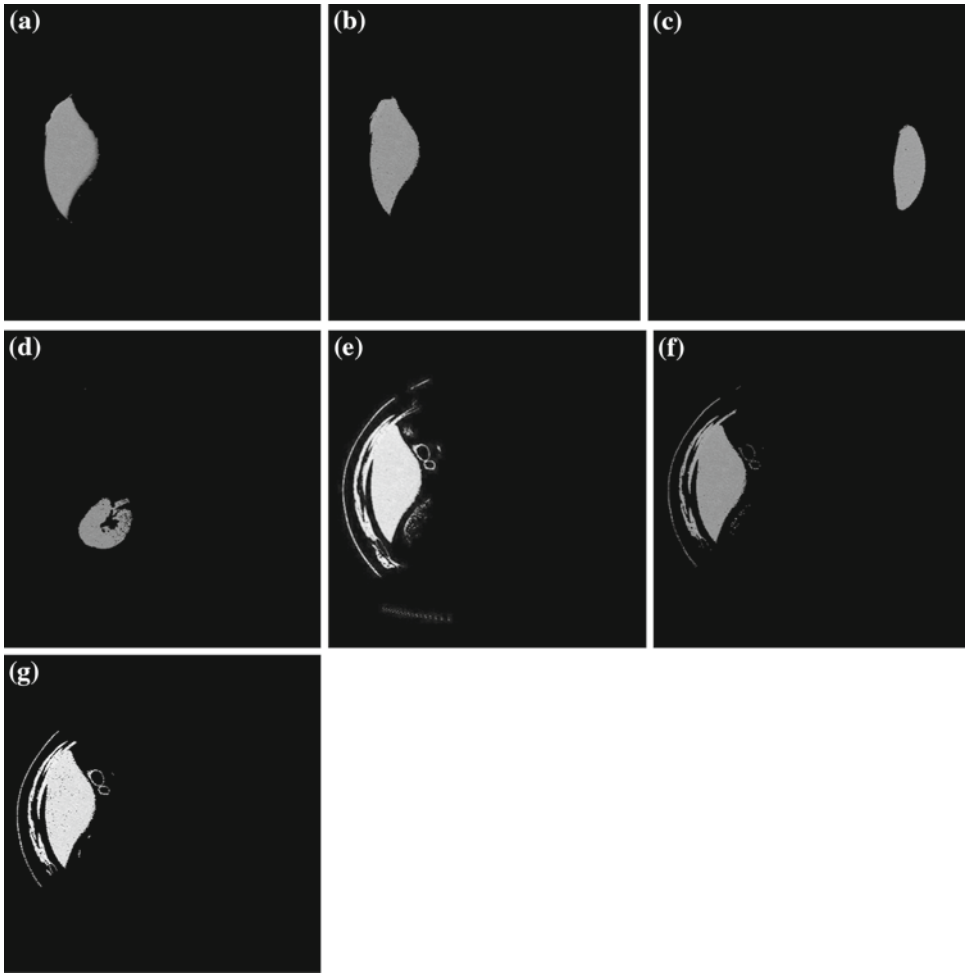

Fig. 11 Estimated tissue types of Fig. 6; **a** Ground truth of liver; **b** liver by our method; **c** spleen by our method; **d** left kidney by our method; **e** liver by F-GMMs; **f** liver by Y-GMMs; **g** liver by C-FCM

3.4 Medical image segmentation with GMMs

The EM algorithm in Sect. 2 and the number of components K are used to estimate the parameter vector θ of SY. According to formula (2), compute the probability $f_r(x_i|\mu_r, \sum_r)$ $(1 \leq r \leq K)$. Each point is classified into the model that has the maximum probability.

4 Experiments

In order to evaluate the performance of the proposed algorithm, we compare it with the Cai's FCM method (C-FCM) of [12], the Figueiredo's GMMs method (F-GMMs) [26] and the Yang's GMMs method(Y-GMMs) [19]. Tests were performed with the simulated MR images [27] with ground truth provided by the Internet Brain Segmentation Repository (IBSR), Center for Morphometric Analysis at Massachusetts General Hospital. The real CT abdomen images from the Affiliated Hospital of Jiangsu University in China were also tested. All experiments were performed on a PC with 1.73 GHz Intel, 1024 MB of RAM. The Dice similarity measure (DSM) [28] is used as performance index and it is defined as:

$$DSM(r) = 2N_{p\cap g}(r)/(N_p(r) + N_g(r)) \tag{24}$$

where $N_{p\cap g}(r)$ denotes the number of pixels classified by both the proposed method and the ground truth as model r, and $N_p(r)$ and $N_g(r)$ represent the number of pixels classified as model r by the proposed method alone and by the ground truth, respectively. The DSM index attains the value 1 if the proposed method coincides with the ground truth and decreases toward 0 as the quality of the segmentation deteriorates. Typically, a value $DSM > 0.7$ means that there is an excellent agreement between the two segmentations [21].

4.1 Validation using digital brain phantoms

The anatomical model used to generate simulated brain MRI data consists of a set of 3d "fuzzy" tissue membership

Table 4 The average DSM index and time of the 5 CT images

	Time(s)	DSM for images		
		Liver	Spleen	Left kidney
C-FCM	50.87	0.73	0.78	0.68
F-GMMs	113.65	0.71	0.65	0.63
Y-GMMs	110.43	0.73	0.72	0.71
Our method	10.15	0.84	0.82	0.75

volumes, one for each tissue class (white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), fat ...). The volumes are defined at a 1 mm isotropic voxel grid in Talairach space, with dimensions $181 \times 217 \times 181$. We choose 10 BrainWeb Images for axial slice no. 66, 71, 76, 81, 86, 91, 96, 101, 106 and 111 to segment by C-FCM, F-GMMs, Y-GMMs and our method. Figure 2 shows some of the 10 images in our tests, while Fig. 3 shows those images with 3% noise. For comparative purposes with the published results, the segmentation was restricted to four components: GM, WM, CSF and "background" for C-FCMs and F-GMMs. Our method and Y-GMMs estimate the number of components automatically. The ground truth of CSF, WM and GM from the BrainWeb images for axial slice no.66 in Fig. 2a are shown in Fig. 4a–c, respectively. The segmentation of CSF, WM and GM from Fig. 2a by our method is shown in Fig. 5a–c, respectively. The segmentation of CSF from Fig. 2a by C-FCM, F-GMMs and Y-GMMs are shown in Fig. 5d–f, respectively.

The DSM indexes for those 20 images of C-FCM, F-GMMs, Y-GMMs and our method are shown in Table 2. We can know that the average DSM of the 10 images by our method and Y-GMMs are higher than the other two methods from Table 2. The main reason is that K in the Y-GMMs and our method is larger than $K = 4$ in the C-FCM and F-GMMs. Our method estimates the number of components $K = 6$, while Y-GMMs estimates $K = 5$.For the 10 images with 3% noise, our method has a higher DSM index than the other methods. The kernel density estimation method and sample technique reduce the effect of the noise data. The time of our method is obviously less than the other three methods. It mainly owes to our histogram stratification sampling, approximate density estimation and the dynamic step size in hill-climbing algorithm.

4.2 Validation by real CT images

We had copied about 16G real CT images from the Affiliated Hospital of Jiangsu University in China for the research. In this paper, we choose 5 real CT abdomen images to segment. Two of those 5 images are shown in Figs. 6 and 7. The ground truths of livers, spleen and kidneys in those 5 images are manually segmented by Tian L.Y, a radiologist of the Affiliated Hospital of Jiangsu University.

The image X of Fig. 6 has N =141,281 points. By algorithm1, the sample image SX has n =19,624 points in Fig 8. From Figs. 6 and 8, we can know that the sample image remains the basic content structure of X. However, the size of SX is only 14% of X. Because the clusters in Fig. 6 are transformed by MDS, the shapes of clusters in Fig. 9 are not identical to the shapes of clusters in Fig. 8. The approximate density function of Fig. 9 is shown in Fig. 10 by Algorithm 2. Each hill in the approximate density function in Fig. 10 is corresponding to one cluster in Fig. 8. Although sampling method may lead the instability of cluster shapes, the number of clusters in Fig. 10 is almost stable. From Table 3, the size of sample n varies from 7,000 to 21,000, but the value of K only varies from 16 to 18. It verifies the stability of our method. The ground truth of liver in Fig. 6 is shown in Fig. 11a, and clusters of liver, spleen, left kidney in Fig. 6 by our method are segmented and shown in Fig. 11b–d, respectively.

Figure 11e–g show the liver of Fig. 5 by F-GMMs, Y-GMMs and C-FCM, respectively. The quantitative results of those four methods are shown in Table 4. From it, we can know that the best advantage of our method is the minimum time cost. The time of our method is only about 10% that of F-GMMs, Y-GMMs and 20% that of C-FCM. All this owes mainly to the histogram stratification sampling, approximate density estimation with the binning strategy and hill-climbing algorithm with the dynamic step size. The DSM indexes for liver, spleen and left kidney by our method are better than F-GMMs, Y-GMMs and C-FCM. The better results of our method should be attributed to our choice of the "true" models. The points of muscle nearby liver should not be in the same cluster of liver. From Fig. 11e–g, the other three methods cannot segment the points of the muscle from the liver. The main reason is that the other three methods are sensitive to the noise data.

5 Conclusion

A fast medical image clustering segmentation algorithm is presented in this paper. The histogram stratification sam-

pling, approximate density estimation with the binning strategy and hill-climbing algorithm with the dynamic step size can improve the speed of segmentation. Our fast strategies make mixture density clustering more practicable to segment medical images with large size. Non-parametric kernel density estimation can estimate the density function of any complex data with only weak general prior assumptions about the data. Therefore, our method with the kernel density function of medical image can accurately estimate the number of components. Moreover, the kernel density estimate is insensitive to noise. In comparison with the common mixture density clustering algorithms and the FCM method for medical image segmentation, the experimental results show that our proposed algorithm has better segmentation results and performance.

Acknowledgments The authors would like to thank the referees for their helpful comments and suggestions to improve the presentation of this paper. This work was supported by the National Science Foundation of China under Grant No. 60841003.

References

1. Zhang, D.Q., Chen, S.C.: A novel kernelized fuzzy C-means algorithm with application in medical image segmentation. Artif. Intell. Med. **32**(1), 37–50 (2004)
2. Bricq, S., Collet, C.H., Armspach, J.P.: Unifying framework for multimodal brain MRI segmentation based on Hidden Markov chains. Med. Image. Anal. **12**(6), 639–652 (2008)
3. Pham, D.L., Xu, C.: Current methods in medical image segmentation. Ann. Biomed. Eng. **2**(8), 315–338 (2000)
4. Zhang, H., Jason, E.F., Sally, A.G.: Image segmentation evaluation: a survey of unsupervised methods. Comput. Vis. Image. Und. **110**(2), 260–280 (2008)
5. Marroquin, J.L., Vemuri, B.C., Botello, S., Calderon, F.: An accurate and efficient Bayesian method for automatic segmentation of brain MRI. IEEE Trans. Med. Imag. **21**(8), 934–945 (2002)
6. Chao, W.H., Chen, Y.Y., Lin, S.H.: Automatic segmentation of magnetic resonance images using a decision tree with spatial information. Comput. Med. Imag. Grap. **33**(2), 111–121 (2009)
7. Iscan, Z., Yüksel, A., Dokur, Z.: Medical image segmentation with transform and moment based features and incremental supervised neural network. Digit. Signal Process **19**(5), 890–901 (2009)
8. Tang, H., Dillenseger, J.L., Bao, X.D., Luo, L.M.: A vectorial image soft segmentation method based on neighborhood weighted Gaussian mixture model. Comput. Med. Imag. Grap. **33**(8), 644–650 (2009)
9. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surveys **31**(3), 264–323 (1999)
10. Kuo, W.F., Lin, C.Y., Sun, Y.N.: Brain MR images segmentation using statistical ratio: Mapping between watershed and competitive Hopfield clustering network algorithms. Comput. Meth. Prog. Bio. **91**(3), 191–198 (2008)
11. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. SIAM Rev. **25**(3), 442–442 (1983)
12. Cai, W.L., Chen, S.C., Zhang, D.Q.: Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Pattern Recogn. **40**(3), 825–838 (2007)
13. Yong, Y.: Image segmentation based on fuzzy clustering with neighborhood information. Opt. Appl. **39**(1), 135–147 (2009)
14. Guillemaud, R., Brady, M.: Estimating the bias field of MR images. IEEE Trans. Med. Imaging **16**(3), 238–251 (1997)
15. Wells, W.M., Grimso, W.E.L., Kikinis, R.: Adaptive segmentation of MRI data. IEEE Trans. Med. Imaging **15**(4), 429–442 (1996)
16. Hayit, G., Adi, T.P.: Medical image categorization and retrieval for PACS using the GMM-KL framework. IEEE T. Inf. Technol. B **11**(2), 190–202 (2007)
17. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum-likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B **39**(1), 1–38 (1977)
18. Tang, Y.G., Liu, D., Guan, X.P.: Multi-resolution image segmentation based on Gaussian mixture model. J. Syst. Eng. Electron **17**(4), 870–874 (2006)
19. Yang, X.Y., Shankar, M.: Image segmentation using finite mixtures and spatial information. Image Vision Comput. **22**(9), 735–745 (2004)
20. Khayati, R., Vafadust, M., Towhidkhah, F.: Fully automatic segmentation of multiple sclerosis lesions in brain MRFLAIR images using adaptive mixtures method and Markov random field model. Comput. Biol. Med. **38**(3), 379–390 (2008)
21. Adelino, R.F.D.S.: A Dirichlet process mixture model for brain MRI tissue classification. Med. Image Anal. **11**(2), 169–182 (2007)
22. McLachlan, G., Peel, D.: Finite Mixture Models. Wiley series in probability and statistics, Chap. 1. pp. 1–20. Wiley, New York (2000)
23. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. **35**(3), 1065–1076 (1962)
24. Govindarajulu, Z.: Elements of Sampling Theory and Methods, Pearson Education, Chap. 5. pp. 75–105 (1999)
25. Laurens, M.V.D., Geoffrey, H.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(1), 2579–2605 (2008)
26. Figueiredo, M.A.T., Jain, A.K.: Unsupervised learning of finite mixture models. IEEE T. Pattern Anal. **24**, 381–396 (2002)
27. http://www.bic.mni.mcgill.ca/brainweb
28. Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C.: Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans. Med. Imaging **13**(4), 716–724 (1994)