A vectorial image soft segmentation method based on neighborhood weighted Gaussian mixture model.

Hui Tang, Jean-Louis Dillenseger, Xu Dong Bao, Li Min Luo

HAL Id: inserm-00411983

https://inserm.hal.science/inserm-00411983

Submitted on 31 Aug 2009

A vectorial image soft segmentation method based on neighborhood weighted Gaussian mixture model

Hui Tang[1,3], Jean-Louis Dillenseger[2,3], Xu Dong Bao[1,3] and Li Min Luo[1,3]

[1]Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University,

210096, Nanjing, China

[2]INSERM U642, Laboratoire Traitement du Signal et de l'Image, Université de Rennes I, 35042 Rennes, France

[3]Centre de Recherche en Information Biomédicale Sino-Français (CRIBs)

Abstract:

The CT uroscan consists of three to four time-spaced acquisitions of the same patient. After registration of these acquisitions, the data forms a volume in which each voxel contains a vector of elements corresponding to the information of the CT uroscan acquisitions. In this paper we will present a segmentation tool in order to differentiate the anatomical structures within the vectorial volume. Because of the partial volume effect (PVE), soft segmentation is better suited because it allows regions or classes to overlap. Gaussian mixture model is often used in statistical classifier to realize soft segmentation problems by getting classes probability distributions. But this model relies only on the intensity distributions, which will lead a misclassification on the boundaries and on inhomogeneous regions with noise. In order to solve this problem, a neighborhood weighted Gaussian mixture model is proposed in this paper. Expectation Maximization algorithm is used as optimization method. The experiments demonstrate that the proposed method can get a better classification result and is less affected by the noise.

Keywords: Gaussian mixture model, vectorial image segmentation, soft segmentation, neighborhood, image classification, 3D/3D registration.

1. Introduction

The CT uroscan is the classical preoperative examination for renal surgery. It consists of three to four time-spaced 3D acquisitions at several contrast medium diffusion stages, which give complementary information about the kidney anatomy. The integration of this information within a unique spatial volume gives the surgeon the knowledge of the patient specific renal anatomy. The first step in this integration process is to bring the different acquisitions into spatial alignment which has been done through a local mutual information maximization registration technique [1]. After registration, the aligned data forms a vectorial volume dataset in which each voxel contains a vector of n elements corresponding to the information of the CT uroscan acquisitions (n is equal to the number of acquisitions, three to four in our case). In order to get the material (tissue) distribution information of this vectorial volume, a multi-dimensional segmentation or classification method should be performed.

Due to partial volume effects (PVE), the voxel intensities at the object boundaries are usually the result of the combination of several materials. Getting the material probabilities by a soft segmentation method [2] instead of assigning a definite material to the voxels (especially the boundary voxels) will be more conformable to the reality.

In the range of segmentation methods, clustering algorithms are termed unsupervised classification methods which organize unlabeled feature vectors into clusters or "natural groups" such that samples within a cluster are more similar to each other than samples belonging to different clusters. The three most commonly used clustering methods are the K-means [3], the fuzzy c-means (FCM) algorithm [4, 5] and the Gaussian mixture model (GMM) [6-8] solved by Expectation Maximization (EM) algorithm [9]. Among the three

methods, fuzzy c-means and Gaussian mixture model have the ability to perform soft segmentation by getting class probability distributions. The fuzzy c-means estimates the parameters which minimize the distance from each voxel to the class centers. It uses only the distance objective function without any other information about the intensity distributions. In contrast, the method based on Gaussian mixture model uses the statistical theory to model each voxel's intensity, which is more reasonable to the real situation. In this paper, we choose the Gaussian mixture model and estimate the Maximum Likelihood parameters by EM algorithm.

Unfortunately, the intensity classification methods rely only on the intensity distributions which will lead to misclassification at the object boundaries. In addition, the lack of information during classification will lead to sensitiveness to noise in inhomogeneous regions. In his tutorial [10] G. Kindlmann noted that for intensity-only classification problems "histograms/scatter-plots entirely loose spatial information" and he asked if there would be "any way to keep some of it?". Many researchers have realized the importance of spatial information for image classification. As described by Roettger et al. [11], spatial information is important, because a feature by definition is a spatially connected region in the volume domain with a unique position and certain statistical properties. These authors indicated that only using the statistical information of the scatter-plot will effectively ignore the most important part of a features definition. Zhang et al. [12] proposed a novel hidden Markov random field (HMRF) model to integrate spatial information to Gaussian model based segmentation methods. Instead of using Markov random field (MRF) as a general prior in Gaussian model based approach as other researchers did [13], the authors proposed a Gaussian hidden Markov

random field model and used a Markov Random Field-Maximum A Posteriori (MRF-MAP) approach to estimate class labels, while MAP was used to estimate the bias field in MR images. Because the bias field doesn't exist in CT images and in addition this model estimates a definite class label for each pixel without the consideration of PVE, this method doesn't meet our requirement. Tang et al. [14] proposed to use a multi-resolution Gaussian mixture model method for image segmentation in order to solve the noise sensitivity problem of Gaussian mixture model based method. This method was realized by constructing an image pyramid which is composed of a sequence of images with reduced resolution. The bottom of the pyramid is the original image and the higher level image represents the lower resolution. According to the pyramid constructing rules, one pixel in the high level of the image pyramid contains larger neighboring information than that in the lower level so that the spatial information is implicitly contained in the higher level of the pyramid. Chuang et al. [4] integrated the spatial information to fuzzy c-means algorithm by incorporating it into the membership function.

In order to integrate spatial information to the Gaussian mixture model based vectorial data segmentation method, we proposed to involve a neighborhood weight within the classification process. To reach this goal, we need a neighborhood information descriptor. Lunstrom et al [15] proposed the Partial Range Histogram (PRH) concept, which is a way to describe the amount of a tissue within a local region. This gives us the hint to use this concept as a neighborhood descriptor. Based on this idea, a neighborhood weighted Gaussian mixture model is proposed in this paper.

2. Gaussian mixture model

If K is the number of tissues (or materials), the Gaussian mixture model assumes that each voxel is composed by K component densities mixed together with K mixing coefficients. Each component density follows a Gaussian distribution. Based on statistical theory, the parameters are estimated by maximum likelihood (ML) and expectation maximization (EM) algorithm is used as an optimization method.

For a vectorial volume with N voxels, the voxel intensity vector is denoted by $x_i (i = 1, 2, \cdots, N)$. Recall that the goal is to estimate the class probabilities on each voxel according to the intensity vectors. The probability distribution of the kth tissue material is denoted by $p_k(x \mid \Theta_k)$, which is governed by a set of parameters Θ_k. Given the parameters of all the classes, the probability distribution of each voxel can be described as a mixture of probability distributions as follows:

$$p(x \mid \Theta) = \sum_{k=1}^{K} \alpha_k p_k(x \mid \Theta_k) \tag{1}$$

where α_k denotes the mixture coefficients. The parameter set of this distribution is $\Theta = (\alpha_1, \cdots, \alpha_K, \Theta_1, \cdots, \Theta_K)$ with the constraint that $\sum_{k=1}^{K} \alpha_k = 1$.

Typically, $p_k(x \mid \Theta_k)$ is modeled by a Gaussian distribution with mean μ_k and covariance matrix Σ_k. That is：

$$
\begin{aligned}
p_k(x \mid \Theta_k) &= p_k(x \mid \mu_k, \Sigma_k) \\
&= \frac{1}{\sqrt{\det(2\pi\Sigma_k)}} e^{-(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)/2}
\end{aligned} \tag{2}
$$

ML estimation is a common used method to find the probability distribution parameters. The log-likelihood expression for this density from the data X is given by:

$$\log(L(\Theta \mid X)) = \log \prod_{i=1}^{N} p(x_i \mid \Theta)$$

$$= \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \alpha_k p_k (x_i \mid \Theta_k) \right)$$

(3)

Finding the ML solution directly from Eq. (3) is difficult because it contains the log of

the sum. The EM algorithm is a good way to solve this problem [16]. The iterative solution

for finding the parameters at the $(t+1)$th iteration step is as follows (M-step):

$$\alpha_k^{t+1} = \frac{1}{N} \sum_{i=1}^{N} p(k \mid x_i, \Theta^t)$$

(4)

$$\mu_k^{t+1} = \frac{\sum_{i=1}^{N} x_i \cdot p(k \mid x_i, \Theta^t)}{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t)}$$

(5)

$$\Sigma_k^{t+1} = \frac{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t) \cdot (x_i - \mu_k^{t+1})(x_i - \mu_k^{t+1})^T}{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t)}$$

(6)

Taking the mixing parameters α_k as prior probabilities, the probability of each class can

be computed using Bayes' rule (E-step):

$$p(k \mid x_i, \Theta^t) = \frac{\alpha_k^t p_k (x_i \mid \Theta_k^t)}{p(x_i \mid \Theta^t)}$$

$$= \frac{\alpha_k^t p_k (x_i \mid \Theta_k^t)}{\sum_{j=1}^{K} \alpha_j^t p_j (x_i \mid \Theta_j^t)}$$

(7)

3. Proposed neighborhood weighted method

The iteration formula described in section 2 didn't involve any spatial information about

current voxel. As discussed in section 1, neighborhood information is one of the most

important spatial information. If the iteration procedure takes the neighborhood effect into

account, the classification result can be more reasonable.

The original model calculates the class probabilities according to Bayes' rule, which is

described by Eq. (7). This calculation is based on intensity distributions without any

neighborhood information. Usually the material is continuous, so that it is natural to have the

idea that for each voxel, the probability of the kth class should be affected by the neighbors' kth class probabilities. According to this belief, we can integrate the neighborhood effect on the class distributions of the current voxel by modifying Eq. (7).

Due to the deducing process of EM algorithm and the natural continuous properties of the classes (materials), the class probability should obey the two rules:

1) $\sum_{k=1}^{K} p(k \mid x_i, \Theta^t) = 1$,

2) Current voxel's kth class probability magnifies if the neighbors' kth class probabilities tend to 1; current voxel's kth class probability decreases if the neighbors' kth class probabilities tend to 0.

According to the second rule, the neighborhood class distribution can be designed as a weight on the current class distribution, so that we designed the neighborhood weighted probability for the current voxel:

$$p(k \mid x_i, \Theta^t) = \frac{\alpha_k^t W_{ik}^t p_k (x_i \mid \Theta_k^t)}{\sum_{j=1}^{K} \alpha_j^t W_{ij}^t p_j (x_i \mid \Theta_j^t)} \tag{8}$$

where

$$W_{ik} = \frac{\sum_{n=1}^{|N_i|} p\left(k \mid x_{ni}, \Theta^t\right)}{|N_i|} \tag{9}$$

N_i is a set of neighborhood of the ith voxel. $|N_i|$ denotes the number of voxels in a set N_i. x_{ni} denotes the nth neighbor's intensity of the ith voxel.

The new class distribution formula is conformable to the two rules (see the prove in the appendix) and integrates the neighborhood information to the current voxel's class distribution during iteration. For each iteration step, the class distribution will be amended by the neighbors' class distribution information. So that through this weighted formula, the

neighborhood information is taken into account to the classification process.

The EM solution formula for the proposed neighborhood weighted Gaussian mixture model is summarized as follows:

E-step:

$$p(k \mid x_i, \Theta^t) = \frac{\alpha_k^t W_{ik}^t p_k(x_i \mid \Theta_k^t)}{\sum_{j=1}^{K} \alpha_j^t W_{ij}^t p_j(x_i \mid \Theta_j^t)} \tag{10}$$

M-step:

$$\alpha_k^{t+1} = \frac{1}{N} \sum_{i=1}^{N} p(k \mid x_i, \Theta^t) \tag{11}$$

$$\mu_k^{t+1} = \frac{\sum_{i=1}^{N} x_i \cdot p(k \mid x_i, \Theta^t)}{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t)} \tag{12}$$

$$\Sigma_k^{t+1} = \frac{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t) \cdot (x_i - \mu_k^{t+1})(x_i - \mu_k^{t+1})^T}{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t)} \tag{13}$$

$$W_{ik}^{t+1} = \frac{\sum_{n=1}^{|N_i|} p\left(k \mid x_{ni}, \Theta^t\right)}{|N_i|} \tag{14}$$

Based on these equations, the estimation process we implemented is described as follows:

(Input: the vectorial volume $x_i (i = 1, 2, \cdots, N)$, the number of classes K.)

Step 1: Initialization of Θ^0, α^0 and W^0. Any classification method could be used, in our case we choose K-means. Using Eq. (9) to initialize the neighborhood weight for each voxel.

Step 2: Calculate the prior probability by Eq. (10).

Step 3: Compute the new parameter data according to Eqs. (11), (12), (13) and (14).

Step 4: Repeat steps 2-3 until reaching the end condition.

For each element vector of the input series, the aim is to find its class distributions. From the iteration process, we can see that this algorithm is not limited neither in applying on

vectorial volume, nor by the spatial dimension of the data. The shape of N_i in Eq. (14) can

match the spatial dimension of any input series and also controls the topology of the spatial

neighborhood information (nearest neighbors or more). The neighborhood shape and topology

could be modified according to the real practical situation.

4. Experiments and discussions

We evaluate our method first on synthetic 2D data and then applied it on real 3D uroscans.

The only difference between vectorial 2D image and 3D vectorial volumes is the shape of

neighborhood accordingly to the data dimension. In order to avoid the local maxima, the

algorithm is initialized as follows: for the synthetic data, we initialize the center points with a

random data near the global maximum; for the practical situation, this initialization can be

implemented by picking a point in each material region manually.

4.1. Evaluation on synthetic data

We choose to evaluate our method on a three-element vectorial image, in which each pixel

contains three intensities. This image is constructed as following (Fig. 1). Each pixel $f(i, j)$ is

composed by three channels[1]: $f(i, j) = [\, f_1(i, j), f_2(i, j), f_3(i, j)]$ and each channel f_1, f_2, f_3 can be

seen as an image. The first channel image f_1 (Fig. 1, left) is composed by two homogeneous

regions in which we add Gaussian noise: ($\mu = 3, \sigma = 1.5$) and ($\mu = 10, \sigma = 1.5$) respectively. The

second channel image f_2 (Fig. 1, middle) follows the same scheme with: ($\mu = 15, \sigma = 1.5$) and

($\mu = 5, \sigma = 1.5$) respectively. The third channel image f_3 (Fig. 1, right) follows also the same

scheme with: ($\mu = 2, \sigma = 1.5$) and ($\mu = 8, \sigma = 1.5$) respectively.

The combination of these three channels leads to a vectorial image with six classes.

---

[1] The word channel is used by analogy with the R, G, B channels of a RGB image.

According to the proposed algorithm described in Section 3, the input number of classes is $K=6$. For the first experiment, $|N_i|$ is set to the 8 closest neighbors.

The classification on synthetic data is performed and the result is shown in Fig. 2. Each pixel of the result image is formed by this formula:

$$C(x_i) = \sum_{k=1}^{K} C_k \, p(k \mid x_i, \Theta) \qquad (15)$$

Where $C(x_i)$ is the color assigned to the ith pixel and C_k is the color we assigned to the kth class.

Fig. 2(a) is the classification result with the original Gaussian mixture model and Fig 2(c) is the result with the neighborhood weighted iterative method. In Fig 2(a) we can notice that the final regions are not homogeneous as expected because of the noise. The reason is that the method relies only on the intensity distribution (histogram). The classification progress is a direct mapping from intensity to classes so that the noise areas are assigned an incorrect class distribution. In Fig. 2(c), it is obvious that the regions are more homogeneous and the classification process is less affected by the noise, when we integrate the neighborhood information into the iteration procedure.

In order to demonstrate that our method is not only a simple class decisions smoothness. We use Eq. (9) to just smooth the class decisions after classification with the classical Gaussian mixture model. The result of the smoothing is illustrated in Fig. 2(b). We can see that although the noise is faded, the edges are blurred at the same time, which is not corresponding to the real situation. This demonstrates the necessity to integrate the neighborhood information within the iteration procedure.

The difference between these results comes from the calculation of $p(k \mid x_i, \Theta)$. Fig. 2(a)

is calculated according to Eq. (7), that is, without any neighborhood weight. Fig. 2(b) and (c) are both calculated according to Eq. (8) with the neighborhood weight W_{ik}, but the apparent results are quite different. If in the conditions with and without neighborhood, the estimated Gaussian mixture parameters are both close to the real parameters (we will demonstrate it later), the difference between Fig. 2(b) and (c) appear on the weight W_{ik}. In Fig. 2(b) it is calculated according to the final result of the class parameters estimation process, in contrast, W_{ik} in Fig. 2(c) is iteratively estimated during the EM algorithm, as presented in section 3. Comparing Fig. 2(b) and (c), we can easily reach the conclusion that when taking the estimation of neighborhood weight into the EM algorithm, the classification result is more conformable to the reality.

The estimated parameters of the Gaussian mixture model should be close to the real parameters which are known for the test data in this paper. In order to demonstrate that integrating the neighborhood weight doesn't reduce the parameter estimation precision. We evaluate the quantitative error of the parameters by the following equation:

$$\varepsilon(r) = \frac{1}{K} \sum_{k} \frac{\left\| r_{k}' - r_{k}^{0} \right\|}{\left\| r_{k}^{0} \right\|} \tag{16}$$

where r_{k}' denotes the estimated parameter and r_{k}^{0} denotes the real parameter value.

The estimation parameters in the condition with and without neighborhood information are listed in Table 1. The estimation error with neighborhood information is a little lower than that without neighbor so that we can reach the result that integrating the neighborhood information into the classification process will not reduce the estimation precision. We use the histogram to visually verify this result because histograms are often considered as an approximation of the class distribution. The classified class distributions should be fitted with

the histograms. It is difficult to illustrate the histogram and the result intensity distribution directly because for three-element vectorial image the histogram has four axes so that we show the histograms by calculating the intensity distribution summation along one axis to reduce the total number of axis. The results are shown in Fig. 3. Fig. 3(a) illustrates the intensity distribution summation along three axes of the original image. Fig. 3(b) is the classification result of Gaussian mixture classification and Fig. 3(c) illustrates the neighborhood weighted classification results. From these figures, we can see that the classification fits the original histogram very well. As the neighborhood weighted method doesn't change the intensity distributions, the effect of our proposed method is to amend the class decision by its neighborhood on each pixel during the iteration procedure, while keeping the global intensity distribution nature.

We also did some experiments while changing the size of the neighborhood in order to see its effect on the classification results. As the real class distribution of the test image is known, we use the misclassification rate to evaluate the classification results. While changing the shape of the neighborhood, the misclassification rate remains around 0.3% (while without neighborhood, 9.2%). The visual results accord with this phenomenon (Fig. 4) so that the shape of the neighborhood plays a tiny part in the classification procedure for this synthetic data. In our case, we choose the 8-closest neighborhood.

4.2. Application on real data

After the evaluation on synthetic data, we performed the methods on the real data obtained after the registration of three CT acquisitions of a human kidney. Fig. 5 shows one slice of the vectorial volume, which is composed by three channels: (a), acquisition before contrast

medium injection; (b), immediately after injection; (c), ten minutes after injection.

We expect to classify the vectorial volume into four classes: fat, renal cortex, renal medulla and collecting system. According to the conclusion of the previous section, we set $|N_i|$ to the 26 closest neighbors. With K=4, the classification result formed by Eq. (15) is shown in Fig. 6. It effectively demonstrates our conjecture. While taking the neighborhood information into classification procedure (Fig. 6(c)), the anatomical structures are better delineated into homogeneous regions: fat (red), renal cortex (green), renal medulla (blue) and collecting system (white).

From the above results, we can reach the conclusion that the Gaussian mixture model based method has the ability to classify vectorial image with the aim of delineating the anatomical structures. Because of the inhomogeneity of the acquisitions and the partial volume effects, the result of the intensity-only method has some misclassification area, especially the renal cortex and the renal medulla because of their close intensity range, which is shown in Fig. 6(a). In order to show this phenomenon more clearly, the corresponding first order derivate of the result probabilities along one cut line (represented in white) is drawn in Fig. 7. Because of the white background color, the collecting system (represented in white originally) is represented in black line in the probabilities derivate figure (Fig. 7), the other materials are represented according to the colors originally assigned to them.

In Fig. 7(a), according to the probabilities first order derivate, we can clearly see that the regions are not separated because there are some inhomogeneous regions, e.g. in the renal medulla (between index [30, 40] on the line) or partial volume effect on tiny object, e.g. the collecting system (between index [70, 80] on the line). While taking the neighborhood

information into the iteration process, the results are improved significantly, as shown in Fig. 7(b). We can see that inside one material region, the proposed method gives a more homogeneous decision. The proposed method considers the intensity and the position of one pixel simultaneously so that it can give a more reasonable classification result. While comparing Fig. 7(a) and (b), we can see that the proposed method has the effect to be less sensitive to inhomogeneities, while giving a better class distribution.

5. Conclusions

In this paper we present a soft segmentation method adapted to classify the information contained in vectorial images or volumes. This method is based on modeling the voxels' intensity vectors distribution by a mixture of Gaussians. We add spatial information on this classical statistical classifier by proposing neighborhood weighted method. In this new model, the voxels' intensity vectors follow still the Gaussian mixture distribution but the classes distributions on each voxel are affected by its neighbors' class probability distributions. The neighborhood information is integrated into the classification process by amending the voxel's class distributions at each iteration step.

Experiments on synthetic and real data have been performed. The results show that this improvement on Gaussian mixture model is less affected by noise and gives better classification results. We also evaluate the impact of the neighborhood. A nearest neighborhood is generally enough in practical use moreover the experiments also prove that the expansion of the neighborhood range makes tiny effect on the estimated Gaussian mixture model parameters. The experimental results demonstrate also that the proposed method gives more reasonable class distributions for each pixel while keeping the global intensity

distribution.

References:

[1]   H. Tang, J. L. Dillenseger, and L. M. Luo, "Intra subject 3D/3D kidney registration using local mutual information maximization," presented at Conf of IEEE EMBS (EMBC'07), Lyon, France, 2007.

[2]   D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, vol. 2, pp. 315-337, 2000.

[3]   G. B. Coleman and H. C. Andrews, "Image segmentation by clustering," *Proceedings of the IEEE*, vol. 67, pp. 773-785, 1979.

[4]   K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol. 30, pp. 9-15, 2006.

[5]   M. S. Yang, Y. J. Hu, K. C. Lin, and C. C. Lin, "Segmentation techniques for tissue differentiation in MRI of ophthalmology using fuzzy clustering algorithms," *Magn Reson Imaging*, vol. 20, pp. 173-9, 2002.

[6]   R. Guillemaud and M. Brady, "Estimating the bias field of MR images," *IEEE Trans Med Imaging*, vol. 16, pp. 238-51, 1997.

[7]   W. M. Wells, W. L. Grimson, R. Kikinis, and F. A. Jolesz, "Statistical intensity correction and segmentation of MRI data," presented at 3rd Conf. Visualization Biomed. Computing (VBC'94), Rochester, MN, 1994.

[8]   W. M. Wells, W. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Trans Med Imaging*, vol. 15, pp. 429-42, 1996.

[9]   J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," International computer science institute, Berkeley, California ICSI-TR-97-021, 1998.

[10]   G. Kindlmann, "IEEE visualization 2001 tutorial 5, presentation 3: transfer functions for direct volume rendering," presented at Visualization'01, San Diego, CA, 2001.

[11]   S. Roettger, M. Bauer, and M. Stamminger, "Spatialized transfer functions," presented at EUROGRAPHICS-IEEE VGTC Symposium on Visualization, Leeds, UK, 2005.

[12]   Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Trans Med Imaging*, vol. 20, pp. 45-57, 2001.

[13]   T. Kapur, E. L. Grimson, R. Kikinis, and W. M. Wells, "Enhanced spatial priors for segmentation of magnetic resonance imagery," presented at MICCAI'98, Cambridge, MA, 1998.

[14]   Y. G. Tang, D. Liu, and X. P. Guan, "Multi-resolution image segmentation based on Gaussian mixture model," *Journal of Systems Engineering and Electronics*, vol. 17, pp. 870-874, 2006.

[15]   C. Lundstrom, P. Ljung, and A. Ynnerman, "Extending and simplifying transfer function design in medical volume rendering using local histograms," presented at EUROGRPHICS - IEEE VGTC Symposium on Visualization, Leeds, UK, 2005.

[16]   A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1-38, 1977.

Appendix

Now we briefly prove that the designed formula Eq. (8) can satisfy the two rules:

1) $\sum_{k=1}^{K} p(k \mid x_i, \Theta^t) = 1$,

2)   Current voxel's kth class probability magnifies if the neighbors' kth class probabilities tend to 1; current voxel's kth class probability decreases if the neighbors' kth class probabilities tend to 0.

When taking Eq. (8) into the left side of the first rule, it equals to 1 so that Eq. (8) can rule 1) naturally.

In order to prove rule 2), we assume that for the kth class of the ith voxel, there exist two neighborhood weights W_1 and W_2 with: $W_1 > W_2 > 0$, the corresponding probabilities calculated by Eq. (8) are denoted by p_1 and p_2 respectively. If we can prove that $p_1 > p_2$, the second rule can be satisfied. We denote $M = \alpha_k^t p_k(x_i \mid \Theta_k^t)$, $N = \sum_{j=1\cdots K, j \neq k} \alpha_j^t W_{ij}^t p_j(x_i \mid \Theta_j^t)$.

When the variable is the neighborhood weight, M and N are constant and $M \geq 0$, $N \geq 0$. We can rewrite Eq. (8) for W_1 and W_2 respectively:

$$p_1 = \frac{MW_1}{MW_1 + N}, \quad p_2 = \frac{MW_2}{MW_2 + N}$$

so that

$$p_1 - p_2 = \frac{MN(W_1 - W_2)}{(MW_1 + N)(MW_2 + N)}$$

When M and N are constant and positive, we can deduce that if $W_1 > W_2 > 0$, then $p_1 > p_2$.

When the condition is $W_{ik} = 0$, the probability $p(k \mid x_i, \Theta^t) = 0$. That is to say, the weighted probability according to Eq. (8) is a monotonously increasing function to the neighborhood weight W_{ik}. With this property, the rule 2) can also be satisfied.