

Multiresolution Fuzzy C-Means Clustering Using Markov Random Field for Image Segmentation

Xuchao Li

College of Information Science and Media, Jinggangshan University, Ji'an, China
bsx7096@sohu.com

Suxuan Bian

College of Nursing, Jinggangshan University, Ji'an, China
lixuchaoliaoning@sina.com

Abstract—In this paper, an unsupervised multiresolution image segmentation algorithm is put forward, which combines interscale and intrascale Markov random field and fuzzy c-means clustering with spatial constraints. In the initial label determination of wavelet coefficient phase, the statistical distribution property of wavelet coefficients is characterized by Gaussian mixture model, the properties of intrascale clustering and interscale persistence of wavelet coefficients are captured by Markov prior probability model. According to maximum a posterior rule, the initial label of wavelet coefficient from coarse to fine scale is determined. In the image segmentation phase, in order to overcome the shortcomings of conventional fuzzy c-means clustering, such as being sensitive to noise and lacking of spatial constraints, we construct the novel fuzzy c-means objective function based on the property of intrascale clustering and interscale persistence of wavelet coefficients, taking advantage of Lagrange multipliers, the improved objective function with spatial constraints is optimized, the final label of wavelet coefficient is determined by iteratively updating the membership degree and cluster centers. The experimental results on real magnetic resonance image and peppers image with noise show that the proposed algorithm obtains much better segmentation results, such as accurately differentiating different regions and being immune to noise.

Index Terms—image segmentation, Markov random field, wavelet transform, fuzzy c-means, multiresolution, scale

I. INTRODUCTION

Segmentation has been found extensive applications in areas such as quantitative analysis of brain tissues, scene analysis, content-based image retrieval and object tracking[1,2]. The purpose of image segmentation is to partition the given image into a number of regions with significantly different properties, such as statistical and structural information. Segmentation is implemented by assigning each pixel to one of a finite number of classes or labels, based on certain contextual property of the neighbouring pixels.

In recent years, image segmentation technique based on stochastic field model has received extensive attention, one of the most influence on image segmentations is Markov random field (MRF)[3]. It is roughly classified into three types, namely, the single resolution and multiresolution segmentation based space domain [4, 5], multiresolution segmentation based wavelet domain [6].

The single resolution segmentation approach assumes that the spatial distribution of regions consists of a single random field, the feature field based on original image is described by conditional probability function given a label field, the label field is modeled as MRF to impose smoothness constraints on the image segmentation [7]. The technique based on single resolution has demonstrated substantial success for image segmentation, but the single resolution MRF model usually favors the formation of large regions and leads to over segmentation, although the segmentation performance can be improved by using a larger neighborhood for each pixel, it maybe blurs the high frequency of image, these shortcomings produce unfavourable influence on image analysis.

The multiresolution segmentation based space domain [8] assumes that the spatial distribution of regions consists of a serial of random field models from coarse to fine scale, the coarse scale label field affects the fine scale label field by the interscale label transition probability of MRF. The approach takes into account the interscale contextual information of pixels, it can improve the accuracy of boundary localization, but when the image is blurred by noise, the spatial domain multiresolution approach will lose the performance of region classification and boundary detection. Although spatial filter can reduce the effect of noise, but the high frequency information of image is badly worn out. Moreover, the feature field model is still established by probability density function (pdf) of pixels, which can not accurately capture the wide range frequency features of image. Since image has locally varying statistical distributions such as texture regions, smooth shading regions, flat regions or their combinations of abrupt features, the spatial non-stationary feature contradicts single frequency statistical characterization based on pixels.

Manuscript received February 13, 2009; revised July 25, 2009; accepted September 15, 2009.
corresponding author: Xuchao Li

The multiresolution segmentation based wavelet domain receives extensive attention. Wavelet transform decomposes an image into a casual hierarchical sub-images, which are characterized by a serial of MRF models with good time-frequency characteristics[9]. The approach can incorporate the interscale persistence and intrascale clustering properties of wavelet coefficients into the label field prior probability model, it can overcome the shortcoming that the label prior probability model easily produces over segmentation. Moreover, wavelet transform can reduce noise without smearing high frequency information of image, since the features of image are extracted by high pass and low pass filter with different frequencies, the filter bank adapt to the non-stationary characterization of image.

In the last decade, soft computing approaches, such as k-means, possibilistic c-means (PCM)[10] and fuzzy c-means (FCM) clustering[11], have been successfully used for image segmentation. K-means forces each pixel to be associated with exactly one label, it is not realistic, since the uncertainty is almost everywhere in medical image, particularly in the case of magnetic resonance image (MRI) due to partial volume effects and noise (during the acquirement). Consequently, the labels of border between tissues can not be thought as belonging to a single region and must be determined by the degree of membership, fuzzy method is a strong tool to solve the problems.

PCM clustering algorithm[10] is firstly proposed by Krishnapuram Raghu and Keller James M. in 1996, it is a variant FCM algorithm, the PCM objective function is constructed by the Euclidean distance between the likelihood function of sample and clustering centers, the algorithm can overcome the unfavourable influence of noise or outliers on image segmentation. But the difficulty of the algorithm is how to determine the likelihood function of characterizing samples and cluster centers, and how to incorporate spatial constraint into the objective function, which directly affects the segmentation results. Moreover, PCM algorithm is sensitive to initialization, it is usually initialized by the standard FCM algorithm.

The first studied FCM algorithm by Jim Bezdek [12], who constructed FCM objective function for image segmentation, but the algorithm has some limitations, firstly which is sensitive to noise, noise can be reduced by the chosen space filter groups, but the edges of image is seriously lost. Secondly, lacks of spatial context information constraints, the spatially neighboring pixels maybe belong to the same region that can be of great aid in image segmentation. Thirdly, is sensitive to initial values and easily converges a local maxima.

In order to overcome the shortcomings of standard FCM clustering and PCM algorithm without taking into account spatial constraints, many researchers have incorporated local spatial statistical information into the conventional clustering algorithm[13] to improve the performance of image segmentation by modifying the objective function, but the improved objective function is established in single or multiresolution space domain[14] instead of wavelet domain.

In this paper, a wavelet domain MRF and FCM clustering algorithm with spatial constraints for robust multiresolution image segmentation is proposed, it has the following features:

The given image is decomposed into different subbands, which can reduce noise and accurately extract the features of image. In order to represent the observed image for multiresolution Bayesian segmentation, we construct a serial of feature field and label field models.

At the initial label determination of wavelet coefficient stage, the feature field is described by Gaussian mixture model (GMM)[15], it can provide a good fit to the global statistical distribution of wavelet coefficients. In order to capture the intrascale clustering and interscale persistence properties of local wavelet coefficients, the contextual statistical information of intrascale and interscale label field models are captured by Markov prior probability model, the model parameters are estimated by improved expectation maximization (EM), according to maximum a posterior (MAP) estimate, the initial label of wavelet coefficient across scales is determined. Taking advantage of the initial label of wavelet coefficient, we can obtain initial clustering centers and the degree of membership for the following improved FCM algorithm.

At the image segmentation stage, we incorporate the local statistical distribution and contextual information of wavelet coefficients into the FCM objective function, the local statistical information captures the clustering property of intrascale wavelet coefficients, the contextual information captures the persistence property of interscale wavelet coefficients. We optimize the modified FCM objective function, an unsupervised multiresolution image segmentation algorithm from coarse to fine scale is obtained.

The rest of this paper is organized as follows: In section 2, two-dimensional discrete wavelet transform (DWT) of image is briefly introduced, then the feature field of image is described by GMM, the intrascale and interscale contextual information is captured by MRF, the model parameters are estimated by improved EM algorithm, according to MAP rule, the initial label of wavelet is determined. Section 3 gives the modified FCM objective function, which incorporates the intrascale clustering and interscale persistence properties of wavelet coefficients, optimizing the new objective function, an unsupervised image segmentation algorithm is obtained. The proposed is outlined in section 4. Simulation results on real MRI and peppers image with noise are given in section 5, and some conclusions are drawn in section 6.

II. THE INITIAL LABEL DETERMINATION OF WAVELET COEFFICIENT

A. Image Multiresolution Representation by DWT

Wavelet transform is a multiresolution analysis technique that has been developed and applied in various fields, such as astronomy, finances, quantum physics, signal processing, video compression and image processing[16,17], etc. Wavelet has a varying window size and finite duration to fit low and high frequency

information of non-stationary signals, and produces an optimal time-frequency resolutions from coarse to fine scale [18].

The continuous wavelet transform of a 1-dimensional signal $f(x)$ is defined as

$$Wf(a,b) = \langle f, \varphi(a,b) \rangle = \int_{\mathbb{R}} f(x) \varphi_{a,b}(x) dx, \quad (1)$$

$$\varphi_{a,b}(x) = |a|^{-\frac{1}{2}} \varphi\left(\frac{x-b}{a}\right). \quad (2)$$

Where, a is called scale parameter that controls wavelet frequency, b is called translation parameter that controls wavelet position (time), $\langle \cdot, \cdot \rangle$ denotes inner product. The function $\varphi_{a,b}$ is called mother wavelet, which satisfies $\int_{\mathbb{R}} \varphi_{a,b}(x) dx = 0$ and generates the other wavelet functions with parameters a dilation and b translation.

The two-dimensional wavelet transform is performed by consecutively convolving one-dimensional filter bank with the rows and columns of image. The given image is decomposed by applying low-pass and high-pass filters associated with a mother wavelet as shown in Fig.1. $H_0(z)$ and $H_1(z)$ denote the one-dimensional low-pass and high-pass filters, respectively, while 1:2 represents downsampling by a factor of 2 in horizontal and vertical direction of image. Fig.1 shows four subband images with $N/2^j \times N/2^j$ resolution in j -level wavelet transform, LL_j denotes approximate subimage, we can recursively decompose it further to obtain next level multiresolution representation of the original image, and three subimages, LH_j , HL_j , HH_j denote the horizontal, vertical and diagonal subband details. Fig. 2(b) shows that the image DWT decomposition and wavelet coefficients distribution. The given image is decomposed by three level wavelet transform in Fig.2(b), as can be seen, the wavelet transform preserves better the spatial-frequency features of image. As far as frequency property of wavelet coefficients is concerned, the stationary portions of image is described by a few large magnitude value wavelet coefficients. Moreover, the magnitude do not vary significantly across scales. On the contrary, the non-stationary portions (around edges) of image is described by large magnitude value wavelet coefficients. As far as spatial property is concerned, the spatially relative position of wavelet coefficients remains unchange from coarse to fine scale. In other words, the persistence of wavelet coefficient magnitude propagates across scales, the property indicates the label of interscale wavelet coefficient maybe remains unchangeable. In Fig. 2(c), the subimages of three level wavelet transform can be represented by hierarchical structure, each coefficient in coarse scales has four wavelet coefficients in the next finer scale, as the pink color box shows the relationship between “parent” and its four “children”. According to the persistence property of interscale wavelet coefficients, it indicates that the label of children is closely similar to its immediate parent. Fig.2(d) shows the statistical distribution of the HH_2 subband wavelet coefficients, it means that intrascale wavelet coefficients have clustering

and compaction properties, the label of each wavelet coefficient depends on its direct neighbouring wavelet coefficients.

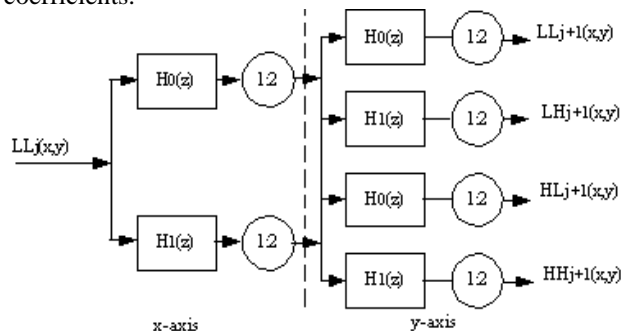
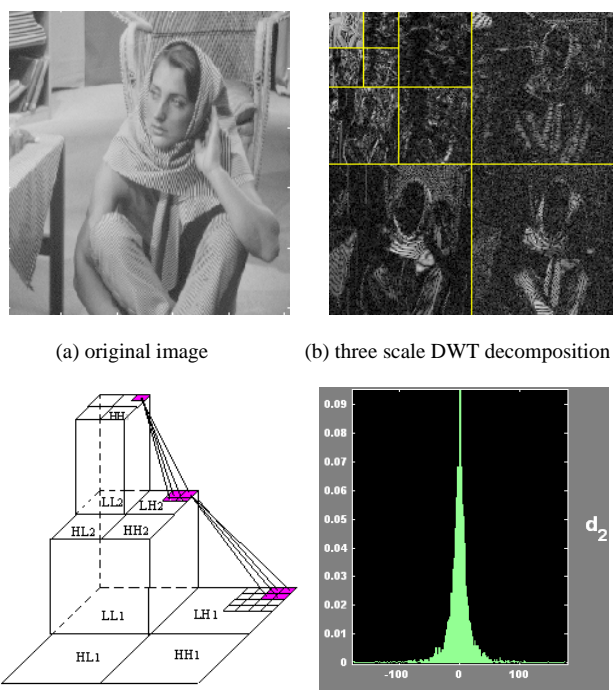


Figure 1. Tree representation j-scale DWT decomposition



(c) subband hierarchical representation (d) coefficients distribution
Figure 2. Image DWT decomposition and coefficients distribution

B. Establishment of Feature Field Statistical Model

After the given image is decomposed by wavelet transform, we are interested in subimages of wavelet domain instead of the spatial domain where the original image exists. Because of multiresolution time-frequency property of wavelet transform, subband wavelet coefficients reflect the statistical features of subimages. Fig. 2(d) displays the histograms of wavelet coefficients of the HH_2 subband of the original image, the wavelet coefficients exhibit approximately Gaussian statistics, GMM[15] can approach arbitrary pdf. To better model the subband statistical features of wavelet coefficients, we utilized GMM for approximating the histograms of wavelet coefficients. According to the decorrelated property of wavelet coefficients, we can assume that coefficients subbands are conditionally independent, the model likelihood of wavelet coefficients is expressed as the following product

$$f(\omega | \theta) = \prod_{k=1}^{N/2^j \times N/2^j} f(\omega_k | \theta_k) \quad (3)$$

Assuming that the image consists of M distinct regions, the model likelihood function of wavelet coefficient is approximated by GMM with M components, it is defined as

$$f(\omega_k | \theta_k) = \sum_{m=1}^M \pi_k^m g(\omega_k, \theta_k^m) \quad (4)$$

Where, π_k^m is the mixture ratio of the m -th component $g(\omega_k, \theta_k^m)$ of GMM and satisfies

$$\sum_{m=1}^M \pi_k^m = 1, 0 \leq \pi_k^m \leq 1, 1 \leq m \leq M \quad (5)$$

The conditional likelihood model of wavelet coefficient given the label can be formulated as following

$$g(\omega_k | l_k^m, \theta_k^m) = \frac{\exp[-\frac{1}{2}(\omega_k - \mu_k^m)\sigma_k^{-2(m)}(\omega_k - \mu_k^m)]}{\sqrt{2\pi}\sigma_k^m} \quad (6)$$

Where, $\theta_k^m = \{\mu_k^m, \sigma_k^{2(m)}, m=1, 2, \dots, M\}$, μ_k^m and $\sigma_k^{2(m)}$ are mean and variance.

Thus the overall subband wavelet coefficients distribution of image is parameterized as follows

$$\theta_j^B = \{\pi_k^m, \mu_k^m, \sigma_k^{2(m)}, k=1, 2, \dots, N/2^j \times N/2^j; m=1, 2, \dots, M\} \quad (7)$$

Where j denotes scale, B denotes wavelet coefficient three subbands.

C. MRF Theory in Wavelet Domain

From (3) and (4), the histogram of wavelet coefficients can be approximated by GMM, but GMM only describes wavelet coefficients frequency statistical property, the spatial interactions between wavelet coefficients is not considered. Wavelet coefficients reflect the spatial-frequency information of image, subband wavelet coefficients from totally different two images maybe have the same frequency histogram, however, the spatial structure information of images is totally different. In order to characterize the spatial structure of image, we exploit MRF to describe the local interactions of wavelet coefficients.

After the image is decomposed by J -level wavelet transform, we obtain $3J+1$ subimages. It is assumed that the subimages are defined on a serial of subband lattices

$$S = \{S^{L_j}, S^{B_{j-1}}, \dots, S^{B_1}\}, S^{B_j} = \{S^{HH_j}, S^{HL_j}, S^{LH_j}\}, S^{HH_j} = N/2^j \times N/2^j, N \times N \text{ represents the size of original image. Let } L \text{ denote the random field associated with the labels of the subimages, it is defined as } L = \{l^{L_j}, l^{B_{j-1}}, \dots, l^{B_1}\},$$

$l^{B_j} = \{l^{HH_j}, l^{HL_j}, l^{LH_j}\}$. Let η denote neighborhood system of sites, it is defined as $\eta = \{\eta^{L_j}, \eta^{B_{j-1}}, \dots, \eta^{B_1}\}$,

$\eta^{B_j} = \{\eta_i^{B_j}, i \in S^{B_j}\}$, where, $\eta_i^{B_j}$ is the set of sites neighboring i , $i \notin \eta_i^{B_j}$ and $i \in \eta_k^{B_j} \Leftrightarrow k \in \eta_i^{B_j}$. A random field L is regarded as an MRF on S with respect to a neighborhood system η , the random field is defined as

$$p(l_i^{B_j} | l_{\{\eta_i^{B_j} - i\}}^{B_j}) = p(l_k^{B_j} | l_k^{B_j}, k \in \eta_i^{B_j}) \quad (8)$$

According to Hammersley-Clifford theorem[19], a given random is an MRF if and only if its probability

distribution $p(l_i^{B_j})$ is a Gibbs distribution, the probability distribution of the label field is formulated as

$$p(l_i^{B_j}) = \frac{1}{Q} \exp[U(l_i^{B_j})] \quad (9)$$

Where, Q is a normalizing constant (partition function),

$U(l_i^{B_j})$ is the energy function having the following form

$$U(l_i^{B_j}) = \sum_{k \in \eta_i^{B_j}} V_k(l_i^{B_j}) \quad (10)$$

D. Establishment of Intrascale and Interscale Label Field Prior Probability Model

As we see from Fig. 2(d), the wavelet coefficients have clustering and compression properties. From Fig. 2(b), we can know the neighbor wavelet coefficients maybe have the same labels, so when determining the label of wavelet coefficients, we must consider the influence of intrascale wavelet coefficient labels. To take intrascale spatial context information into account, we adopt MRF prior probability model for imposing the interactions of wavelet coefficient label field. The prior probability model should express the belief that the pixels within a region have a higher probability. We use the second-order neighbourhood system with respect to the Euclidian distance, and only consider the interactions of wavelet coefficients with two pair site clique, as is shown in Fig.3.

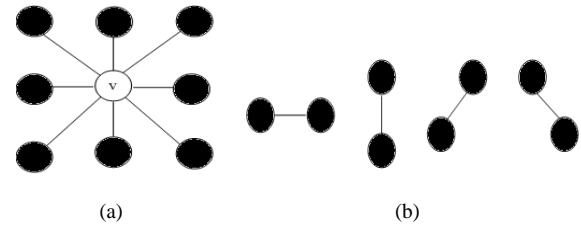


Figure 3. Intrascale neighbourhood and clique. (a) A second-order neighbourhood system with respect to the Euclidian distance. (b) The four associated two-site cliques.

According to (9), the label field prior probability model of intrascale wavelet coefficient is defined as

$$p(l_v) = [\exp(U(l_v))]/Q_1 \quad (11)$$

Here, Q_1 is a normalizing constant called the partition function, v denotes the position of sublattices. $U(l_v)$ is called the intrascale energy function, which has the following form:

$$U(l_v) = \sum_{v \in S^{B_j}} \sum_{\tau \in \eta(v)} V_\tau(l_v) \quad (12)$$

Where $v \in S^{B_j}$, $\eta(v)$ is the second order neighbour system, $V_\tau(l_v)$, which only depends on $l_v, \tau \in \eta(v)$, are the clique potential function. The function has the following form

$$V_\tau(l_v) = -\beta \sum_{v \in S^{B_j}} \sum_{\tau \in \eta(v)} [\delta(l_v - l_\tau) - 1] \quad (13)$$

Where, the Markov parameter β controls the spatial interactions of local wavelet coefficients, so that two neighbouring wavelet coefficients are more likely to have the same label than the different labels. $\delta(\square)$ is the discrete delta function, it is defined as

$$\delta(l_a - l_b) = \begin{cases} 0 & \text{if } (l_a = l_b) \\ 1 & \text{if } (l_a \neq l_b) \end{cases} \quad (14)$$

Fig. 2(b) indicates the interscale wavelet coefficients have persistence property. Fig. 4 shows a wavelet coefficient is decomposed by 3-level wavelet transform, each wavelet in coarse scale has four children in next fine scale. If we want to determine the label of node v at j -th scale, according to the interscale persistence property, the label of its parent node 5 in $(j-1)$ th scale will affect the label of node v , moreover, node 5 has three brothers, which derive from the same parent node 1, their labels will affect the label of node v . We call the nodes of 2,3,4 and 5 as the interscale neighbor of node v . It is assumed that the labels of interscale wavelet coefficients have one order Markov property, The prior probability model of interscale label field with quadtree neighborhood system is defined as

$$p(l_v^{B_j} | l_\gamma^{B_{j-1}}) = \frac{1}{Q_2} \exp[-U_2(l_v^{B_j})] \quad (15)$$

Where, $U_2(l_v^{B_j})$ is called the interscale energy function, it is defined as

$$U_2(l_v^{B_j}) = -\alpha \sum_{v \in s^{B_j}} \sum_{\gamma \in \rho(s^{B_{j-1}})} [\delta(l_v^{B_j} - l_\gamma^{B_{j-1}}) - 1] \quad (16)$$

where α denotes the Markov parameter, $\rho(s^{B_{j-1}})$ denotes the interscale neighbour clique.

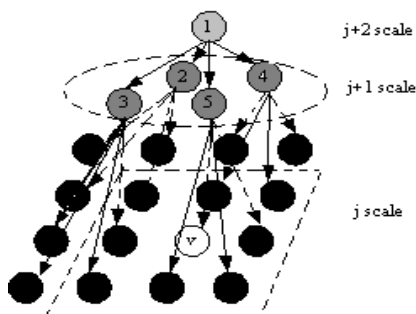


Figure 4. Interscale neighbourhood and clique

E. MAP Estimate of Initial Label

The purpose of multiresolution image segmentation is to assign a class label from the set of L for each subband wavelet coefficient. Let \hat{l} denote the optimum estimate for true but unknown label l^* of wavelet coefficient, \hat{l} and l^* are realization of MRF, which are formulated as Markov label field prior probability model. The observed subband wavelet coefficients ω are formulated as a realization of GMM, which is formulated as feature field conditional probability model given the label field. According to MAP, the optimum criterion of label is formulated as

$$\hat{l} = \arg \max_{l \in L} p(l | \omega, \theta) \quad (17)$$

Where θ is the model parameter.

Using Bayesian rule, equation (17) is reformulated as

$$\hat{l} = \arg \max_{l \in L} p(l | \omega, \theta) p(l) \quad (18)$$

From (18), we need to compute the prior probability of label field and the likelihood probability of feature field. Submitting (6), (11) and (15) into (18), according to MAP equivalent to minimizing energy function, (18) can be expressed as

$$U(l | \omega) = \arg \min_{l \in L} [U(\omega | l) + U_1(l) + U_2(l)] \quad (19)$$

Where $U(l | \omega)$ denotes the posterior energy function of label field.

Submitting (6), (13) and (16) into (19), according to the local property of MRF, the posterior energy function is expressed as

$$U(l_v | l_{\eta(v)}, \omega) \propto \sum_{\tau \in \eta(v), \gamma \in \rho(s^{B_j})} \left\{ \frac{1}{2} [\ln \sigma_k + \frac{(\omega_k - \mu_k)^2}{\sigma_k^2}] + \beta [\delta(l_v - l_\tau) - 1] + \alpha [\delta(l_v - l_\gamma) - 1] \right\} \quad (20)$$

Some stochastic iterative optimization algorithms such as simulated annealing and deterministic relaxation are adopted to get the global minimum of (20), but finding the optimum labels of the posterior energy function requires a lot of computation time. In this paper, we adopt the iterative conditional modes (ICM) algorithm[19] to update the label of wavelet coefficient. Given the wavelet coefficient ω and the initial label of coarse scale wavelet coefficient, the ICM algorithm sequentially updates label $l_v^{(t+1)}$ by minimizing the following formula

$$\hat{l}_v^{(t+1)} = \arg \min U(l_v | l_{\eta(v)}, \omega) \quad (21)$$

Where t denotes the number of iterations.

F. EM Algorithm for Model Parameters Estimation

To implement image segmentation, the parameters θ in GMM have to be estimated, the procedure for determining the unknown parameters is known as model fitting, the objective of model fitting is to obtain the unbiased estimation of model parameters. However, because of the nonlinear relationship between the parameters of GMM and their pdfs in (4) and (6), it is impossible to derive an explicit expression for the maximum likelihood (ML) estimation of parameters. Many iterative algorithms have been employed to solve the problem, among which the EM algorithm[20] is the one most widely utilized.

The classic EM algorithm is a general ML estimation algorithm, it has a well-known basic structure for dealing with the incomplete data, its concrete structure is problem dependent. Here, the determination of wavelet coefficient label is formulated as an incomplete data processing problem. The observed wavelet coefficient ω is considered as the incomplete data characterized by likelihood $p(\omega | \theta)$, let the label l denote unobservable or missing data, let the couple (ω, l) denote the complete data characterized by the joint likelihood function $p(\omega, l | \theta)$, where θ is parameter to be estimated. The ML estimation of θ is determined by the following formula

$$\hat{\theta} = \arg \max_{\theta} \log p(\omega, l | \theta) \quad (22)$$

EM algorithm starts with assigning initial values to the parameter θ , it then iteratively alternates between the expectation step (E-step) and maximization step (M-step)

for approximating the ML estimation of parameter θ . It can be expressed as follows:

E-step: computing the log-likelihood function conditional expectation of the complete data z , given the observed wavelet coefficient ω and the current parameter $\theta^{(t)}$ ML estimation

$$\Phi(\theta | \theta^{(t)}) = E_t[\ln p(\omega, l | \theta) | \omega, \theta^{(t)}] \quad (23)$$

Where $E_t[\square]$ denotes the mathe expectation of random variable l , which is described by the conditional pdf of missing data given the wavelet coefficients.

M-step: the parameter θ is reestimated by maximizing

$$\theta^{(t+1)} = \arg \max_{\theta} \Phi(\theta | \theta^{(t)}) \quad (24)$$

Where t denotes the number of iterations.

From (24), we can get ML estimation of parameter, equation (23) indicates that the ML estimation of parameter is seriously dependent on initialization, consequently, clssical EM algorithm usually obtains the local maxima. Besides, because EM algorithm is a general framework, we must integrate concert application issue into the general framework for obtaining the optimum estimation of parameter. To improve the performance of parameter estimation, we introduce histogram of image as constraint conditions of EM algorithm.

In order to impose constraints, we introduce a hidden variable z , let $z = \{z_p = [z_{pl}, l = 1, 2, \dots, L], p = 1, 2, \dots, P\}$, where z_{pl} denotes the probability of grey level p belongs to the l th region, P denotes the number of grey levels. The improved EM algorithm computes the conditional expectation of the complete data (ω, z) as follows

$$\Phi(\theta | \theta^{(t)}) = \int_z p(z | \omega, \theta^{(t)}) \bullet \ln p(\omega, z | \theta) dz \quad (25)$$

From (25), we do not have any prior knowledge about missing variable z , in order to compute the conditional expectation of complete data, we must obtain the prior probability of missing data z via maximum entropy principle. The maximizing entropy of prior probability of missing data z is defined as

$$H = - \int_z p(z | \omega, \theta) \bullet \ln p(z | \omega, \theta) dz \quad (26)$$

Under the constraint

$$\int_z p(z | \omega, \theta) dz = 1 \quad (27)$$

Equation (27) indicates the constraint must conform to the normalising conditon.

According to Lagrangian theorem, the optimization is to maximize the following objective function

$$E = \Phi(\theta | \theta^{(t)}) + \Gamma H + \lambda \left(\int_z p(z | \omega, \theta) dz - 1 \right) \quad (28)$$

Where Γ and λ are Lagrange multipliers.

By maximizing (28) with respect to the prior probability $p(z | \omega, \theta)$, we have

$$p(z) = \frac{1}{Q_3} [p(\omega, z | \theta)]^\Gamma \quad (29)$$

Where Q_3 is the normalising function that satisfies

$$Q_3 = \int_z [p(\omega, z | \theta)]^\Gamma dz \quad (30)$$

In (29), we introduce a parameter Γ , which is similar to the temperature parameter of simulated annealing

algorithm, during the cooling process, we can get the global optima. The improved EM algorithm is formulated as:

E-step: the conditional expectation of complete data (ω, z) is

$$\Phi(\theta | \theta^{(t)}) = \sum_z \frac{[p(\omega, z | \theta)]^\Gamma}{\sum_z [p(\omega, z | \theta)]^\Gamma} \bullet \ln p(\omega, z | \theta) \quad (31)$$

M-step: the parameter θ is reestimated by maximizing (31).

III. FUZZY CLUSTERING ALGORITHM WITH SPATIAL CONSTRAINTS

According to (21), we obtain the initial labels of wavelet coefficients from coarse to fine scale, then the initial cluster centers of regions are determined. In order to overcome the standard FCM algorithm without considering wavelet domain spatial-frequency property shortcomings[21], we incorporate intrascale local spatial clustering property and interscale frequency persistence property into FCM objective function. The improved FCM objective function is optimized by Lagrange multipliers, we can obtain a new image segmentation algorithm by iteratively updating clustering centers and degree of membership. Taking advantage of maximizing degree of membership criterion, the final segmentation results of original image are obtained.

A. Standard FCM Algorithm

The FCM algorithm assigns each wavelet coefficient to a region label by using fuzzy degree of membership. Let $\omega = \{\omega_{s^j}, j = 1, 2, \dots, J\}$, $\omega_{s^j} = \{\omega_{s^j}^1, \omega_{s^j}^2, \dots, \omega_{s^j}^M\}$, $\omega_{s^j} \in \omega$, denote the j th scale wavelet coefficients to be partitioned into K classes, and the grey value $c = \{c_{s^j}, j = 1, 2, \dots, J\}$, $c_{s^j} = \{c_{s^j}^1, c_{s^j}^2, \dots, c_{s^j}^K\}$, $c_{s^j} \in c$, denote a set of the region cluster centers. In order to achieve optimum clustering, the formulation of the FCM optimization model to be minimized is defined as follows

$$J_m(u, c) = \sum_{i \in s^j} \sum_{l=1}^K (u_{s^j}^{i,l})^m \| \omega_{s^j}^i - c_{s^j}^l \|^2 \quad (32)$$

with the following constraints

$$\sum_{l=1}^L u_{s^j}^{i,l} = 1, 0 \leq u_{s^j}^{i,l} \leq 1 \quad (33)$$

where $u_{s^j}^{i,l}$ denotes the degree of membership of wavelet coefficient $\omega_{s^j}^i$ in the l th cluster, $c_{s^j}^l$ is the l th class center, $\|\bullet\|$ represents the Euclidean distance, and m is a fuzzifizer that controls the resulting partition, $m \in [1, \infty)$.

B. The Intrascale FCM Objective Function

In standard FCM algorithm, the degree of membership is dependent solely on the distance between the wavelet coefficient and each individual cluster center in wavelet domain, and disregards the spatial property of wavelet coefficients. According to the spatial clustering property of intrascale wavelet coefficients, the local neighboring

wavelet coefficients usually possess similar features, the probability that they belong to the same region is great. In order to capture the spatial clustering property, the intrascale FCM objective function is defined as

$$J_{N(\omega_{s^j}^i),m}(u,c) = \frac{1}{N_R} \sum_{i \in N(\omega_{s^j}^i)} \sum_{l=1}^K (u_{s^j}^{i,l})^m \left\| \omega_{N(\omega_{s^j}^i)}^j - c_{s^j}^l \right\|^2 \quad (34)$$

where $N(\omega_{s^j}^i)$ stands for the set of neighboring wavelet coefficients around $\omega_{s^j}^i$, N_R stands for the size of window, and the other parameters are of the same meaning as those in (32).

C. The Interscale FCM Objective Function

Wavelet transform has spatial-frequency property, in (34), we incorporate the local spatial clustering property into the FCM objective function, Fig.2(b) shows that wavelet coefficient has the interscale similar frequency persistence property, the label of child in fine scale is similar to that of its parent's neighbors in coarse scale. In order to capture frequency persistence property, the interscale FCM objective function is defined as

$$J_{\rho(\omega_{s^{j-1}}^i),m}(u,c) = \sum_{i \in \rho(\omega_{s^{j-1}}^i)} \sum_{l=1}^L (u_{i,l})^m \left\| \omega_{\rho(\omega_{s^{j-1}}^i)}^j - c_{s^j}^l \right\|^2 \quad (35)$$

Where $\rho(\omega_{s^{j-1}}^i)$ denotes the set of neighbors of $\omega_{s^{j-1}}^i$ in the $(j-1)$ th scale, $\omega_{s^{j-1}}^i$ is the immediate parent of $\omega_{s^j}^i$ in the j th scale, $\omega_{\rho(\omega_{s^{j-1}}^i)}^j$ denotes the neighboring wavelet coefficients of $\omega_{s^{j-1}}^i$ in the $(j-1)$ th scale, and the other parameters are of the same meaning as those in (32).

D. Local Statistical Fuzzy Degree of Membership

The degree of membership in standard FCM algorithm only depends on the distance between wavelet coefficient and cluster center. In this paper, we incorporate the interscale persistence and intrascale clustering property of wavelet coefficient into the standard FCM objective function, consequently, the degree of membership of new FCM algorithm must change accordingly. According to the spatial-frequency feature of wavelet transform, the intrascale clustering and interscale persistence property of wavelet coefficients is incorporated into the degree of membership, which is defined as

$$(u_{s^j}^i)^* = u_{s^j}^i p_{[N\rho(\omega_{s^j}^i)]}(\omega_{s^j}^i) \quad (36)$$

Under the following constraint

$$0 \leq (u_{s^j}^i)^* \leq 1, \sum (u_{s^j}^i)^* = 1 \quad (37)$$

Where $N\rho(\omega_{s^j}^i) = N(\omega_{s^j}^i) + \rho(\omega_{s^{j-1}}^i)$ denotes the intrascale two order neighbors and interscale parent neighbors of wavelet coefficient $\omega_{s^j}^i$, the intrascale spatial and interscale frequency statistical information $p_{[N\rho(\omega_{s^j}^i)]}(\omega_{s^j}^i)$ is expressed as

$$p_{[N\rho(\omega_{s^j}^i)]}(\omega_{s^j}^i) = \frac{(A+B) \cdot p(\omega_{s^j}^i | l_n)}{\sum_{\omega_{s^j}^n \in N(\omega_{s^j}^i)} p(\omega_{s^j}^n | l_n) + \sum_{\omega_{s^{j-1}}^n \in \rho(\omega_{s^j}^i)} p(\omega_{s^{j-1}}^n | l_n)} \quad (38)$$

where, $\delta(\square)$ is sample function, $A = \sum_{\omega_{s^j}^n \in N(\omega_{s^j}^i)} \delta(l_i - l_n) p(\omega_{s^j}^n | l_n)$,

$$B = \sum_{\omega_{s^{j-1}}^n \in \rho(\omega_{s^j}^i)} \delta(l_i - l_n) p(\omega_{s^{j-1}}^n | l_n) .$$

E. Improved FCM Objective Function

The standard FCM algorithm in (32) does not impose any spatial-frequency constrain on objective function, which can lead to forming small undesirable regions. The objective functions in (34) and (35) take into account the intrascale clustering and interscale persistence property of wavelet coefficient, respectively. Submitting (36) into (32), (34) and (35), the modified FCM objective function in wavelet domain can be written as

$$J_M^*(U^*, C) = J_m(u^*, c) + \lambda_1 J_{N(\omega_{s^j}^i),m}(u^*, c) + \lambda_2 J_{\rho(\omega_{s^{j-1}}^i),m}(u^*, c) \quad (39)$$

Where the parameter λ_1 is a weight that controls the intrascale spatial context information, λ_2 controls the interscale persistence information.

F. Optimization Improved FCM Objective Function

The segmentation algorithm is formulated as minimizing (39) under condition (37), the optimization of (39) will be solved by using Lagrange multiplier technique

$$J_M^*(U^*, C) = J_m(u^*, c) + \lambda_1 J_{N(\omega_{s^j}^i),m}(u^*, c) + \lambda_2 J_{\rho(\omega_{s^{j-1}}^i),m}(u^*, c) + \lambda_3 \left(1 - \sum_{l=1}^K u_{s^j}^{*(i,l)}\right) \quad (40)$$

The derivative of (40) with respect to $u_{s^j}^{*(i,l)}$ and $c_{s^j}^l$ respectively. The modified fuzzy c-means objective function of $J_M^*(U, C)$ can be minimized iteratively by updating membership degree and clustering centroids equation respectively

$$u_{k,i}^* = \frac{\left[\left\| \omega_{s^j}^i - c_{s^j}^l \right\|^2 + \lambda_1 \left\| \omega_{N(\omega_{s^j}^i)}^j - c_{s^j}^l \right\|^2 + \lambda_2 \left\| \omega_{\rho(\omega_{s^{j-1}}^i)}^j - c_{s^j}^l \right\|^2 \right]^{\frac{1}{m-1}} p_{N\rho(\omega_{s^j}^i)}}{\left[\sum_{l=1}^K \left(\left\| \omega_{s^j}^i - c_{s^j}^l \right\|^2 + \lambda_1 \left\| \omega_{N(\omega_{s^j}^i)}^j - c_{s^j}^l \right\|^2 + \lambda_2 \left\| \omega_{\rho(\omega_{s^{j-1}}^i)}^j - c_{s^j}^l \right\|^2 \right) \right]^{\frac{1}{m-1}}} \quad (41)$$

$$c_{s^j}^l = \frac{\sum_{i=1}^n u_{k,i}^* [\omega_{s^j}^i + \lambda_1 \omega_{N(\omega_{s^j}^i)}^j + \lambda_2 \omega_{\rho(\omega_{s^{j-1}}^i)}^j]}{(1 + \lambda_1 + \lambda_2) \sum_{i=1}^n u_{k,i}^*} \quad (42)$$

After obtaining the modified FCM clustering algorithm, each pixel will be associated with a membership degree for each region. By assigning each pixel to the region with the highest membership value, the segmentation result of the image can be obtained.

IV. OVERVIEW OF THE PROPOSED ALGORITHM

The complete procedures from the optimum label establishment to the final modified FCM segmentation algorithm can be stated as follows

A. Determining the label of wavelet coefficients procedure

(1). Feature extraction: performing J level wavelet decomposition of the input image.

(2). Estimating GMM parameters by improved EM algorithm.

(3). According to Bayesian rule, calculating the label of wavelet coefficient using (21).

(4). Given the label of wavelet coefficient, we can get the conditional probability $p(\omega_{s'}^j | I_n)$. Submitting $p(\omega_{s'}^j | I_n)$ into (38), we can calculate the wavelet coefficient spatially statistical information of $P_{N\rho(\omega_{s'}^j)}$.

B. The modified FCM segmentation algorithm procedure

(1). According to the results of the j -th scale wavelet coefficient label estimation, setting the number of clusters, initializing the fuzzy membership function $u_{k,l}^{j(0)}$, $u_{k,l}^{j(0)} \in [0,1]$.

(2). Computing membership functions using (41).

(3). Calculating the fuzzy cluster centers using (42).

(4). if $\|u_{k,i}^{*(\tau+1)} - u_{k,i}^{*(\tau)}\| \leq \varepsilon$ stop; otherwise go to step

B(2). Here, τ represents the number of iterations.

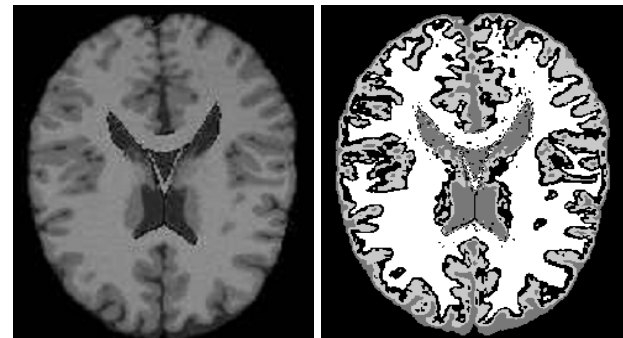
V. SIMULATION RESULTS

We have tested the proposed algorithm via two different images, one is real MRI image, another is a peppers image with noise. In order to evaluate the performance of image segmentation, we compare our simulation results with that of the other segmentation algorithms, standard FCM algorithm and FGMM-MRF model based on single resolution.

Fig.5 shows the segmentation results by applying three different algorithms to MRI image. Fig.5(b) shows the segmentation results using standard FCM algorithm, it produces a lot of isolated regions, because standard FCM algorithm does not take into account the spatial context information of pixels. Fig.5(c) shows the result using FGMM-MRF algorithm, the segmentation result is better than that of FCM algorithm, because the FGMM-MRF algorithm captures the local pixel interactions by MRF prior probability. However, it still produces a few isolated tissues, because the feature model is dependent on single resolution pixels of space domain, which is very difficult to describe the nonstationary property of image. Fig.5(d) shows the segmentation results using the proposed algorithm, the result is better than the others, such as producing few isolated issues, the accuracy of classifying different tissues and edge location.

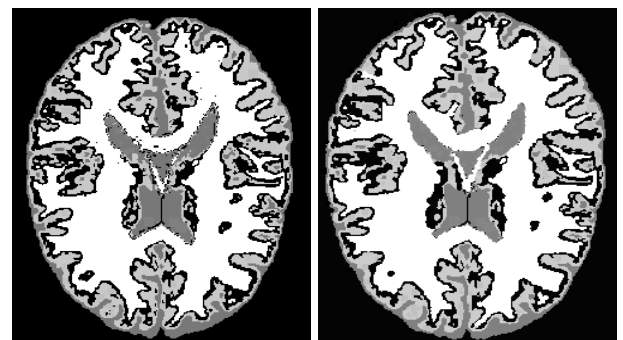
Fig.6 shows the segmentation results by applying three different algorithms to peppers image with noise. Fig.6(b) shows that FCM algorithm is sensitive to noise, the regions are misclassified seriously. Fig.6(c) shows that FGMM-MRF algorithm based on space domain is still sensitive to noise, and produces a lot of isolated regions. Fig.6(d) shows that the proposed algorithm based on wavelet domain is insensitive to noise, the segmentation

result is encouraging, such as reliable edge location and the accuracy of differentiating different regions.



(a) MRI image

(b) FCM segmentation



(c) FGMM-MRF

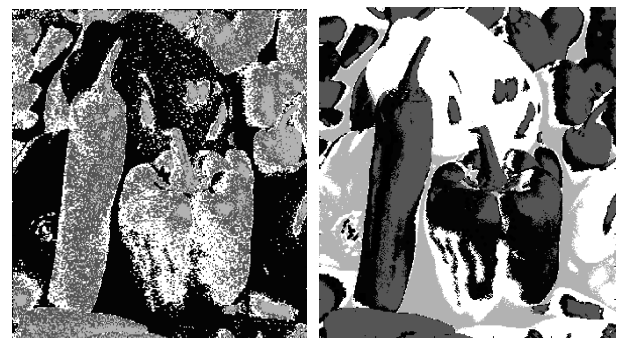
(d) the proposed method

Figure 5. Comparison of segmentation results on MRI



(a) MRI image

(b) FCM segmentation



(c) FGMM-MRF

(d) the proposed method

Figure 6. Comparison of segmentation results on peppers image

VI. CONCLUSION

In this paper, a new multiresolution technique for image segmentation based on wavelet domain spatial-frequency MRF and modified FCM algorithm with interscale and intrascale context constraints is proposed. The multiresolution image segmentation algorithm contains the determination of initial label of wavelet coefficient from coarse to fine scale and image segmentation steps. At the determination of initial label stage, the statistical distribution of feature field in wavelet domain is captured by GMM, the interscale and intrascale label field prior probability models are characterized by MRF, taking advantage of Bayesian rule, the initial label of wavelet coefficient from coarse to fine scale is determined. At the image segmentation stage, a modified FCM objective function is established, it incorporates intrascale clustering and interscale persistence properties into the standard FCM objective function, optimizing the new objective function, an iterative image segmentation algorithm is obtained. Experimental results show that the proposed algorithm is more efficient than standard FCM and FGMM-MRF algorithm, such as being insusceptible to noise, reliable edge location and the accuracy of differentiating different regions.

ACKNOWLEDGMENT

This research work is supported by Natural Science Foundation of Jingtangshan University (No. JZ0822); Youth Natural Science Foundation of Jiangxi Provincial Department of Education, China (No. GJJ09589).

REFERENCES

- [1] Kalyani Mali, Sushmita Mitra, and Tinku Acharya, "A multiresolution fuzzy clustering of images," *International journal of computational cognition*. vol. 4, pp. 30-38, 2006.
- [2] Nuno Vasconcelos, and Andrew Lippman, "Empirical Bayesian motion segmentation," *IEEE Transactions on pattern analysis and machine intelligence*. vol. 23, pp. 217-221, 2001.
- [3] Huawu Deng, David A. Clausi, "Unsupervised image segmentation using a simple MRF model with a new implementation scheme," *Pattern recognition*. vol. 37, pp. 2323-2335, 2004.
- [4] Xuchao Li, Shan'an Zhu, "FGMM-MRF hierarchical model to image segmentation," *Journal of computer aided design & computer graphics*. vol. 17, pp. 2659-2664, 2005.
- [5] Jean-Marc Laferte, Patrick Perez, and Fabric Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Transactions on image processing*. vol. 9, pp. 390-404, 2000.
- [6] J. Sun, D. Gu, S. Zhang and Y. Chen, "Hidden Markov bayesian texture segmentation using complex wavelet transform," *IEE Proceedings on Vision Image Signal Process*. vol. 151, pp. 215-223, 2004.
- [7] M. L. Comer and E. J. Delp, "The EM/MPM algorithm for segmentation of textured images: analysis and further experimental results," *IEEE Transactions on image processing*. vol. 9, pp.1731-1744, 2000.
- [8] Bouman C. A, Shapiro M. "A multiscale random field model for Bayesian image segmentation," *IEEE Transactions on image processing*. vol. 9, pp. 162-177, 1994.
- [9] Justin K. Romberg, Hyeokho Choi, and Richard G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on image processing*. vol. 10, pp. 1056-1068, 2001.
- [10] Krishnapuram Raghu, and Keller James M., "The possibilistic C-means algorithm: insights and recommendations," *IEEE Transactions on fuzzy systems*. vol. 4, pp. 385-393, 1996.
- [11] Keh-Shih Chuang, Hong-Long Tzeng, Sharon Chen, Jay Wu, Tzong-Jer Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*. vol. 30, pp. 9-15, 2006.
- [12] Bezdek Jim, "Pattern recognition with fuzzy objective function algorithm," Plenum, New York, 1981.
- [13] Dzung L. Pham, "Spatial models for fuzzy clustering," *Computer Vision and Image understanding*. vol. 84, pp. 285-297, 2001.
- [14] Mahmoud Ramza Rezaee, Pieter M. J. van der Zwet, Boudewijn P. F. Lelieveldt, Johan H. C. Reiber "A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering," *IEEE Transaction on image processing*. vol. 9, pp. 1238-1248, 2000.
- [15] Zoran Zivkovic, Ferdinand van der Heijden, "Recursive unsupervised learning of finite mixture models," *IEEE Transactions on pattern analysis and machine intelligence*. vol. 26, pp. 651-656, 2005.
- [16] Matthew S. Crouse, Robert D. Nowak, and Richard G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Transactions on signal processing*. vol. 46, pp. 886-902, 1998.
- [17] M.T. Hanna, S.A. Mansoori, "The discrete time wavelet transform: its discrete time fourier transform and filter bank implementation," *IEEE Transactions on Circuits Systems II: Analog Digital Signal Process*. vol. 48, pp. 180-183, 2001.
- [18] Daubechies I, "Ten lectures on wavelets," Philadelphia: Society for Industrial and Applied Mathematics, 1992.
- [19] Djamel Boukerroui, Olivier Basset, Nicole Guerin, Attila Baskurt, "Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation," *European Journal of Ultrasound*. vol. 8, pp. 135-144, 1998.
- [20] A. P. Dempster, N. M. Lard, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *Journal Royal Statistics Society*. vol. 39, pp. 1-37, 1977.
- [21] S. Mallat, S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on pattern analysis and machine intelligence*. vol. 14, pp. 710-732, 1992.

Xuchao Li was born in China in 1974. He received the Bachelor and Master degree in electronic engineering from Liaoning Project Technology University, Fuxin, China, in 2000 and 2003, respectively, and Ph. D degree in control theory and control engineering from Zhejiang University, Hangzhou, China, in 2006. His research was in the area of computer vision and image processing, including object detection and tracking, image denoising, and medical image segmentation.