Texture Segmentation Using Graph Cuts in Spectral Decomposition Based Riemannian Multi-Scale Nonlinear Structure Tensor Space

Shoudong Han and Xinyu Wang

***Abstract*—This paper proposes an interactive texture segmentation method base on graph cuts. It extracts the texture features by using multi-scale nonlinear structure tensor, and discusses dissimilarity measure and probability distribution of features in the Riemannian space, which are used to design the edge-based and region-based items of the segmentation model respectively. To construct distributions, we employ the Gaussian mixture model with covariant-scale based full covariance structure. Additionally, we propose the spectral decomposition based recursive clustering algorithm to estimate the corresponding statistics. The comparisons of various texture segmentation experiments demonstrate the validity of the proposed method.**

***Index Terms*—Multi-scale nonlinear structure tensor (MSNST), graph cuts, texture segmentation.**

I. Introduction

Texture segmentation is a key issue in the field of computer vision. To model the texture features, there are many different approaches, such as multiple resolution techniques, Gabor wavelet filters and so on [1]. However, they all need to estimate many unknown parameters and may include some redundant information. To extract texture features more compactly, the multi-scale nonlinear structure tensor (MSNST) is used in this paper, which has been introduced in our previous work [2].

After the extraction of texture features, the problem is transformed into the matter of how to segment in this new feature space. The graph cuts model proposed in [3] is one of the most widely researched and applied interactive image segmentation methods [4]-[6], and Lazy Snapping [7] and GrabCut [8] are the most successful applications of it. In the literature of [2], it integrates MSNST into the GrabCut framework for color-texture segmentation and obtains the satisfied experimental results. However, [2] discusses the distance measure and probability distribution of MSNST features in the space of information theory, and it does not take into account the Riemannian manifold structure of MSNST. Besides, the Gaussian mixture model (GMM) used in [2] is designed with variance structure, which is not accurate enough when used to describe those feature sets with high relevance. Furthermore, the clustering algorithm adopted in [2] is the simplest K-means clustering, which is seriously sensitive to initial clustering centers. To address the problems mentioned above, this paper proposes a graph cuts based interactive texture segmentation method, and the main idea is: calculating *t-links* base on our new proposed spectral decomposition recursive clustering algorithm and GMM by means of Lazy Snapping type interactions; discussing the distance measure and GMM statistics of MSNST features in the Riemannian manifold space, and designing the GMM with covariant-scale based full covariance structure.

Manuscript received April 2, 2014; revised June 24, 2014. This work was supported in part by the National Natural Science Foundation of China (Grant 61105006), the Research Fund for the Doctoral Program of Higher Education (Grant 20110142120047), and Fundamental Research Funds for the Central Universities of China (HUST: 2013QN150).

S. D. Han is with the National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: shoudonghan@hust.edu.cn).

X. Y. Wang is with the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: wxyhust@163.com).

II. Multi-Scale Nonlinear Structure Tensor

To construct the MSNST, we need first to define the multi-scale structure tensor (MSST), and then apply the nonlinear filtering for smoothing.

MSST can be obtained by using the non-orthogonal (redundant) discrete wavelet frames [9]. Let $\theta(x, y)$ be a 2-D differentiable smoothing function. Define two wavelets $\psi^x(x,y) = \partial\theta(x,y)/\partial x$ and $\psi^y(x,y) = \partial\theta(x,y)/\partial y$. Let

$$\begin{cases} \psi_s^x(x,y) = \alpha^{-s}\psi^x\left(\alpha^{-s}x, \alpha^{-s}y\right) \\ \psi_s^y(x,y) = \alpha^{-s}\psi^y\left(\alpha^{-s}x, \alpha^{-s}y\right) \end{cases},$$

where subscript s denotes the s-th scale, and α can be set as 2. The wavelet transform of the image $I(x, y)$ at the s-th scale has two components, which are named as $D_s^x(x,y) = I * \psi_s^x(x,y)$ and $D_s^y(x,y) = I * \psi_s^y(x,y)$. One can easily prove that

$$\begin{pmatrix} D_s^x(x,y) \\ D_s^y(x,y) \end{pmatrix} = \alpha^s \begin{pmatrix} \partial(I*\theta_s)(x,y)/\partial x \\ \partial(I*\theta_s)(x,y)/\partial y \end{pmatrix} = \alpha^s \nabla(I*\theta_s)(x,y).$$

Therefore, MSST can be constructed using the tensor product of the gradient of $(I*\theta_s)(x,y)$ at each scale as:

$$\mathbf{T}_s = \alpha^{-2s} \sum_{h=1}^{H} \begin{pmatrix} (D_{h,s}^x)^2 & (D_{h,s}^x D_{h,s}^y) \\ (D_{h,s}^x D_{h,s}^y) & (D_{h,s}^y)^2 \end{pmatrix} \quad (1)$$

where $s \in [0, S-1]$, S is the total number of scales, the subscript h denotes the h-th color channel of image I, and H is the total number of the color channels.

Finally, nonlinear diffusion [10] is applied to all scales of MSST separately for computing the MSNST as

$$\partial_t u_{s,i} = \mathrm{div}(g(\sum_{j=1}^{J}\left|\nabla u_{s,j}\right|^2)\nabla u_{s,i}) \quad \forall i$$

where $u_{s,i}(t=0)$ are the independent channels of $\mathbf{T}_s$, $J \equiv 3$ is the total number of independent channels, and g is a decreasing function that can be given as

$$g(|\nabla u|) = \left(|\nabla u|^p + \varepsilon\right)^{-1} \tag{1}$$

Above, ε is a small positive constant, and the constant p is used to balance edge enhancement and smoothing.

To get good texture features, each pixel is represented with the MSNST $\mathbf{\Gamma} = \left\{\hat{\mathbf{T}}_0, \hat{\mathbf{T}}_1, \ldots, \hat{\mathbf{T}}_{S-1}\right\}$, which is described as a set of matrices (they are sized as 2×2). Here, the "hat" denotes that the corresponding component has been nonlinearly diffused, and we will remove the expression of "hat" in the following description for convenience.

III. Texture Segmentation Model

A. Graph Cuts Framework

This paper constructs the MSNST field segmentation model based on graph cuts framework [3], which is posed as the minimizing of the following energy function:

$$E(\alpha) = \sum_{i\in I} D_i(\alpha) + \gamma \sum_{(m,n)\in C \text{ and } \alpha_m \neq \alpha_n} S_{m,n}(\alpha) \tag{2}$$

where α denotes the assigned label, with 0 for background and 1 for foreground, I represents all the image pixels, C is the set of pairs of neighboring pixels, $\gamma \geq 0$ is the smoothness coefficient.

The data term $D_i(\alpha)$ indicates the MSNST similarity of the pixel i to the foreground or background as:

$$D_i(\alpha) = -\ln P\left(\mathbf{\Gamma}_i \middle| \alpha_i\right) \tag{3}$$

Additionally, the smoothness term $S_{m,n}(\alpha)$ evaluates the penalty for assigning neighboring pixels m and n to different regions, can use the MSNST contrast $dis(\cdot)$ as:

$$S_{m,n}(\alpha) = \tau + \left\|m-n\right\|^{-1} \exp\left(-\xi dis^2(\mathbf{\Gamma}_m, \mathbf{\Gamma}_n)\right) \tag{4}$$

where τ is denoising constant, $\|\cdot\|$ is L_2 Euclidean norm, and the parameter ξ can be adaptively set to be

$$\xi = \left(2|C|^{-1} \sum_{(m,n)\in C} dis^2(\mathbf{\Gamma}_m, \mathbf{\Gamma}_n)\right)^{-1}$$

where $|C|$ denotes the number of pairs in the set C, and $dis(\cdot)$ can be defined as the square root of the square sum of tensor distance for all scales as:

$$dis(\mathbf{\Gamma}_m, \mathbf{\Gamma}_n) = \sqrt{\sum_{s=0}^{S-1} dis^2(\mathbf{T}_{m,s}, \mathbf{T}_{n,s})} \tag{5}$$

B. Gaussian Mixed Model (GMM)

In order to fit the formula (4) into the *t-links* of graph cuts framework, we need to define the probability density function (PDF) of extracted features. The usual modeling approaches include thresholding, histogram, extremum picking, nonparametric nearest-neighbor model, support vector machine, Gaussian distribution model, and Markov random field, etc [11]. In this paper, we choose the GMM as the MSNST data modeling, which can be used to describe any complex probability distribution more flexibly and more accurately by means of the linear combinations of some Gaussian distributions. Moreover, in order to maximize the discriminating ability of GMM, the covariant-scale based full covariance structure is employed for it.

Corresponding to the $P(\cdot)$ in (4), the K-component GMM distribution can be described as:

$$P(\mathbf{\Gamma}|\theta) = \sum_{k=1}^{K} \omega_k P(\mathbf{\Gamma}|\theta_k) \tag{7}$$

where $\omega_k \geq 0$ is the mixture weighting coefficient and it meets $1 = \sum_{k=1}^{K} \omega_k$; θ_k and $\theta \equiv \{\omega_1, \ldots, \omega_K, \theta_1, \ldots, \theta_K\}$ are the set of statistics for the k-th component and the whole GMM respectively.

In details, θ_k can be further expressed as $\{\mu_k, \Sigma_k\}$, namely the mean and covariance matrix of MSNST in the covariant-scale space. Here, $\mu_k = \left\{\mu_{k,0}, \mu_{k,1}, \ldots, \mu_{k,S-1}\right\}$ and $\Sigma_k = \frac{1}{|\Omega_k|}\int_{\Omega_k} \varphi'(\mathbf{\Gamma}_x, \mu_k)\ \varphi'(\mathbf{\Gamma}_x, \mu_k)^{\mathrm{T}} dx$, where Ω_k is the cluster region that corresponds to the k-th component. Besides, $\varphi'(\mathbf{\Gamma}, \mu_k) = \left[\varphi(\mathbf{T}_0, \mu_{k,0})^{\mathrm{T}} \quad \cdots \quad \varphi(\mathbf{T}_{S-1}, \mu_{k,S-1})^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\varphi(\mathbf{T}_s, \mu_{k,s})$ is the gradient mapping of tensor manifold vector $\overrightarrow{\mu_{k,s}\mathbf{T}_s}$, and $\mu_{k,s}$ is the s-th scale mean of the MSNST features in Ω_k . So, we can define $P(\mathbf{\Gamma}|\theta_k)$ as:

$$P(\mathbf{\Gamma}|\theta_k) = \frac{\exp\left(-\frac{1}{2}\varphi'(\mathbf{\Gamma}, \mu_k)^{\mathrm{T}} \Sigma_k^{-1} \varphi'(\mathbf{\Gamma}, \mu_k)\right)}{\sqrt{(2\pi)^{3S}|\Sigma_k|}} \tag{8}$$

C. Spectral Decomposition Based Recursive Clustering

In order to estimate the statistics θ in formula (7), the traditional idea is: first of all, given the number of clusters K; and then, initialize the model parameters randomly; finally, update these statistics iteratively using K-means or EM algorithm. However, many problems exist in this kind of clustering. For instance, cluster number K must be given in advance, which does not meet the requirement of unsupervised parameter estimation; the clustering results are usually sensitive to the initializations; since the introduction of iterations, it is bound to reduce the computational efficiency.

To address the problems just mentioned, a recursive

clustering algorithm base on spectral decomposition in the covariant-scale MSNST space is proposed in this paper, which can estimate the statistics directly without any initializations. The main idea is to perform the following recursive process until the termination condition is met: calculate the covariant-scale covariance matrixes of all the current clusters, and select the cluster with the maximum eigenvalue; and then, divide this cluster into two clusters using the spectral decomposition algorithm as shown in Fig. 1. The termination condition here is reaching some specified cluster number *K*.

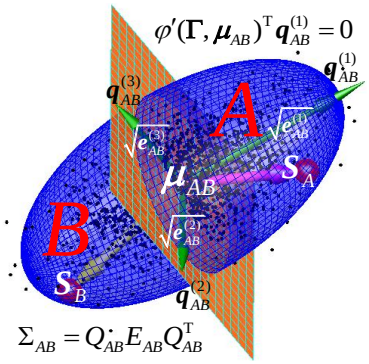


Fig. 1. Spectral decomposition based clustering algorithm.

Assuming ***AB*** is the current selected cluster with the maximum eigenvalue, and here we will give the details of dividing it into cluster ***A*** and cluster ***B*** using the spectral decomposition. By means of the covariant-scale statistical calculations, we can obtain μ_{AB} and Σ_{AB} of the MSNST features in cluster ***AB***. Since Σ_{AB} is symmetric positive semi-definite matrix, we can perform the singular value decomposition as $\Sigma_{AB}=Q_{AB}E_{AB}Q_{AB}^{\mathrm{T}}$. Here, E_{AB} is a diagonal matrix composed of all the eigenvalues of Σ_{AB} in descending order, and the column vectors of Q_{AB} is the eigenvectors corresponding to these eigenvalues. Thus, a splitting hyper-plane equation $\varphi'(\Gamma,\boldsymbol{\mu}_{AB})^{\mathrm{T}}\boldsymbol{q}_{AB}^{(1)}=0$ can be constructed in the covariant-scale MSNST space, where $\boldsymbol{q}_{AB}^{(1)}$ is the eigenvector corresponding to the maximum eigenvalue $\boldsymbol{e}_{AB}^{(1)}$ of E_{AB}. Therefore, for any a random MSNST sample ***S*** in the cluster ***AB***, it can be clustered into cluster ***A*** or ***B*** as:

$$\boldsymbol{S}\in\begin{cases}\boldsymbol{A} & \text{if } \varphi'(\boldsymbol{S},\boldsymbol{\mu}_{AB})^{\mathrm{T}}\boldsymbol{q}_{AB}^{(1)}>0\\ \boldsymbol{B} & \text{default}\end{cases}$$

D. Riemannian Manifold Calculation

Taking into account the positive definite structure of tensor space, the Riemannian manifold calculations of the distance and statistics are more suitable for the MSNST features. According to the definition in [12], the tensor distance for some scale corresponding to formula (6) can be described as:

$$dis(\mathbf{T}_{m,s},\mathbf{T}_{n,s})=\sqrt{\frac{1}{2}\mathrm{tr}\left(\log^2\left(\mathbf{T}_{m,s}^{-1/2}\mathbf{T}_{n,s}\mathbf{T}_{m,s}^{-1/2}\right)\right)}$$

where $tr(\cdot)$ denotes the matrix trace operator.

Algorithm 1: Riemannian estimation of tensor mean.

input: $\{\mathbf{T}_{i,s}\}\in Sym_2^+,\ i=1,2,\ldots,X$

output: μ_s

$(\mu_s)_0=\mathbf{T}_{1,s}$

$V_0=Inf$

$\tau=1,\ \ell=0$

Do

$\ell=\ell+1$

$U_\ell=\frac{1}{X}\sum_{i=1}^{X}\log\left((\mu_s)_{\ell-1}^{-1/2}\mathbf{T}_{i,s}(\mu_s)_{\ell-1}^{-1/2}\right)$

$V_\ell=(\mu_s)_{\ell-1}^{1/2}U_\ell(\mu_s)_{\ell-1}^{1/2}$

$(\mu_s)_\ell=(\mu_s)_{\ell-1}^{1/2}\exp(\tau U_\ell)(\mu_s)_{\ell-1}^{1/2}$

if $\|V_\ell\|>\|V_{\ell-1}\|$

$\tau=\tau/2,\ V_\ell=V_{\ell-1}$

end if

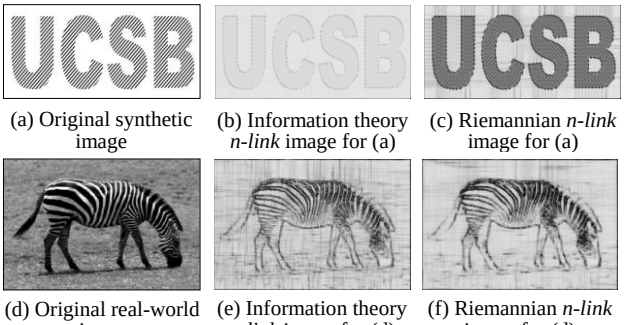
While $\|V_\ell\|>\zeta$

The tensor mean corresponding to formula (8) can be estimated using algorithm 1. Here, X indicates the total number of features, Sym_2^+ is the real symmetric positive definite matrix space sized as 2×2, *Inf* indicates infinite value, and ζ is set as 0.01 in this paper.

The tensor gradient mapping mentioned in (8) can be constructed as $\varphi(\mathbf{T}_s,\mu_s)=\mathrm{Vec}\left(\log\left(\mu_s^{-1/2}\mathbf{T}_s\mu_s^{-1/2}\right)\right)$, where the mapping $\mathrm{Vec}(\mathbf{T})$ equals to $(t_{1,1},\sqrt{2}t_{1,2},t_{2,2})^{\mathrm{T}}$, and it is used to compose the three separate components $t_{1,1}$, $t_{1,2}$ and $t_{2,2}$ of 2×2 sized symmetric matrix $\mathbf{T}$ into a standard orthogonal basis.

IV. Experiment Results and Analysis

In this section, some experiments are given to verify the performance of our proposed method when compared with the literature [2]. There are a number of parameters that must be appropriately determined before. Such as, the number of scales of the MSNST *S* is chosen as 2; the parameters ε and *p* in (1) are fixed as 0.001 and 0.6 respectively; the weighted coefficient γ in (2) is fixed as 50; denoising constant τ in (4) is set as 2.5; in (7), the number of Gaussian components *K* in GMM is set as 5.

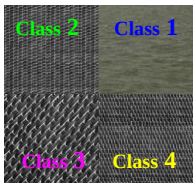


(a) Original synthetic image (b) Information theory *n-link* image for (a) (c) Riemannian *n-link* image for (a)

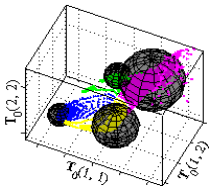
(d) Original real-world image (e) Information theory *n-link* image for (d) (f) Riemannian *n-link* image for (d)

Fig. 2. The compared *n-link* images using the different distance measure in MSNST space.

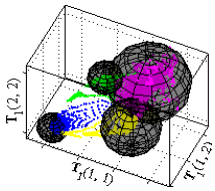
In the first experiment, the compared MSNST *n-link* images between information theory distance (used in [2]) and the Riemannian distance (used in this paper) are visualized in Fig. 2. Since the *n-links* of graph cuts model are mainly determined by the distance measure, we can sum up the 8-neighborhood *n-links* of each pixel for the visualizations. The comparison results verify that, when compared with information theory distance, Riemannian distance obtains smoother regions and larger differences among these regions. Therefore, the Riemannian measure has better descriptiveness about MSNST consistency and the better MSNST discriminating ability.



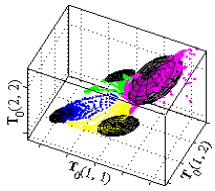
(a) Original image



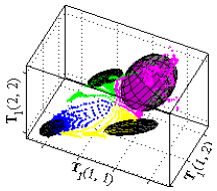
(b) The 0-th scale variance



(c) The 1-th scale variance

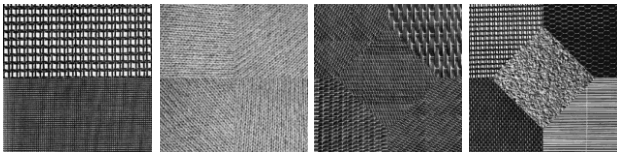


(d) The 0-th scale covariance

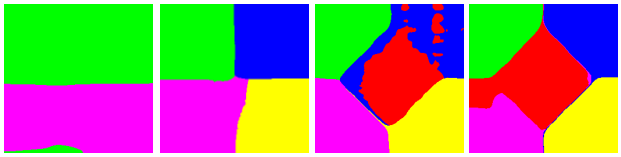


(e) The 1-th scale covariance

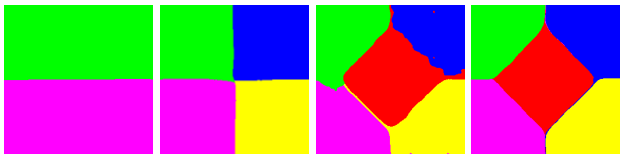
Fig. 3. The compared visualizations of MSNST distributions using the different GMM structures.



(a) The original images



(b) The K-means clustering results of (a)



(c) The spectral decomposition clustering results of (a)

Fig. 4. The compared clustering results between K-means algorithm and spectral decomposition algorithm.

Modeling GMM in the Riemannian manifold space, Fig. 3 further compares the visualizations of MSNST distributions between variance structure (used in [2]) and covariant-scale covariance structure (used in this paper). The comparisons demonstrate that the variance structure can not model the MSNST distributions accurately at any scale. The main reason is that, in practical applications, the distributions of similar features are usually irregular and that of different features are usually much overlapped. Obviously, the isotropy based variance structure is not so flexible and accurate as the anisotropy based covariance structure. The modeling results show that the covariant-scale covariance structure is more suitable to describe the actual distributions of MSNST features.

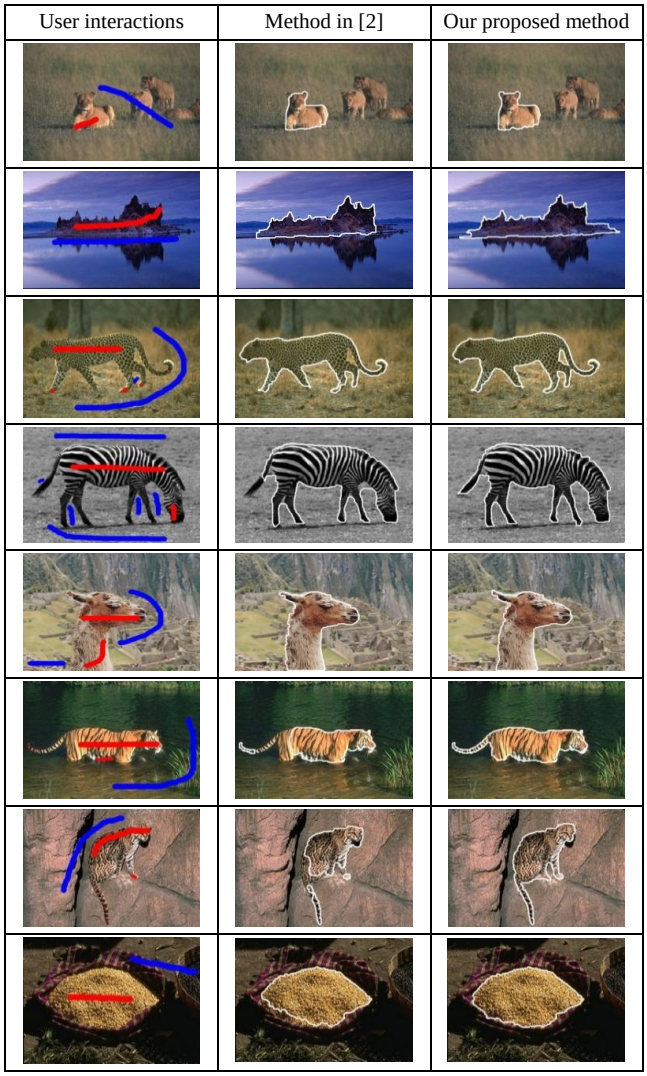


Fig. 5. The compared segmentation results between method in [2] and our proposed method by means of Lazy Snapping type interactions.

According to the conclusions drawn from the above experiments, Fig. 4 compares the texture clustering results between K-means algorithm and spectral decomposition algorithm in the covariant-scale Riemannian space. All the textures used to synthesize the first three experimental images are of the same type, and only differ in orientation or scale, or both, and the last image is synthesized using different kinds of textures. Since the K-means algorithm is seriously sensitive to the initial clustering centers, the displayed clustering result is the best among the 6 random initialization tests. All the experimental results indicate that, no matter what kinds of the texture differences are, the spectral decomposition algorithm is more accurate and more robust than the K-means algorithm.

In order to further demonstrate the superiority of our proposed method, a number of experimental comparisons are shown in Fig. 5 using the GMM based Lazy Snapping framework. The comparison experiments verify that, when segmenting the complex natural texture images base on the same user interactions, our proposed method can perform better than the method in [2] in terms of accuracy,

availability, and practicality.

V. Conclusion

Base on our previous work [2], this paper proposes a more effective interactive texture segmentation method, which uses the GMM based Lazy Snapping framework to complete the texture segmentation, and takes full account of the Riemannian manifold structure of MSNST features to design the distance measure and GMM statistics. Here, the GMM is constructed with covariant-scale covariance structure, and a spectral decomposition based recursive clustering algorithm is proposed to estimate the statistics. The comparison experiments using a large number of synthesized texture images and real natural scene images show that our proposed method obtains more satisfactory segmentation results than that in [2].

References

[1] H. Zhou, J. Zheng, and L. Wei, "Texture aware image segmentation using graph cuts and active contours," *Pattern Recognition*, vol. 46, no. 6, pp. 1719-1733, 2013.

[2] S. D. Han, W. B. Tao, D. S. Wang, X. C. Tai, and X. L. Wu, "Image segmentation based on GrabCut framework integrating multi-scale nonlinear structure tensor," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2289-2302, Oct. 2009.

[3] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 2, pp. 109-131, 2006.

[4] B. Peng, L. Zhang, and D. Zhang, "A survey of graph theoretical approaches to image segmentation," *Pattern Recognition*, vol. 46, no. 3, pp. 1020-1038, 2013.

[5] T. N. A. Nguyen, J. Cai, J. Zhang, and J. Zheng, "Robust interactive image segmentation using convex active contours," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3734-3743, 2012.

[6] W. Yang, J. Cai, J. Zheng, and J. Luo, "User-friendly interactive image segmentation through unified combinatorial user inputs," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2470-2479, 2010.

[7] Y. Li, J. Sun, C. K. Tang, and H. Y. Shum, "Lazy snapping," in *Proc. SIGGRAPH Conf.*, 2004, pp. 303-308.

[8] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: interactive foreground extraction using iterated graph cuts," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 309-314, 2004.

[9] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd, New York: Academic Press, 1999, pp. 259-272.

[10] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 221-232, 1992.

[11] H. Permuter, J. Francos, and I. Jermyn, "A study of Gaussian mixture models of color and texture features for image classification and segmentation," *Pattern Recognition*, vol. 39, no. 4, pp. 695-706, 2006.

[12] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41-66, 2006.

Shoudong Han received the B.S. degree, M.S. degree, and Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2005, 2007, and 2010 respectively.

He is currently an associate professor of School of Automation with HUST since 2013. His research interests lie in the area of computer vision, image segmentation, texture analysis, and pattern recognition. He has published numerous periodical papers and conference papers in these fields, such as the IEEE Trans. Image Process., Pattern Recognition, Pattern Recognition Letters, etc.

Xinyu Wang was born in Wuhan, China, on August 12, 1987. She received the B.S. degree in the Department of Electronics and Information Engineering, Huazhong University of Science and Technology (HUST), Wuhan, China, in 2009. She is currently pursuing the M.S. degree also in the Department of Electronics and Information Engineering at HUST.

Her research interests include computer vision, image processing, and vehicle recognition system.