Pattern Recognition 43 (2010) 2630–2645

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Scale selection for anisotropic diffusion filter by Markov random field model

Jian Sun, Zongben Xu *

Institute for Information and System Sciences, School of Science, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Article history:
Received 30 October 2007
Received in revised form
18 November 2008
Accepted 23 February 2010

Keywords:
Scale selection
Anisotropic diffusion filter
Scale space
Markov random field model
Tree-reweighted message passing
Image denoising
Noise estimation

ABSTRACT

The selection of stopping time (i.e., scale) significantly affects the performance of anisotropic diffusion filter for image denoising. This paper designs a Markov random field (MRF) scale selection model, which selects scales for image segments, then the denoised image is the composition of segments at their optimal scales in the scale space. Firstly, statistics-based scale selection criteria are proposed for image segments. Then we design a scale selection energy function in the MRF framework by considering the scale coherence between neighboring segments. A segment-based noise estimation algorithm is also developed to estimate the noise statistics efficiently. Experiments show that the performance of MRF scale selection model is much better than the previous global scale selection schemes. Combined with this scale selection model, the anisotropic diffusion filter is comparable to or even outperform the state-of-the-art denoising methods in performance.

1. Introduction

Scale space method has been widely investigated in the field of image filtering [1], edge detection [2–4], feature detection [5–7] and data mining [8]. Diffusion filter is an important scale space method widely used in image denoising. Taking noisy image as initial condition, diffusion filter gradually removes the noises and details from the noisy image. The diffused images with increasing scales constitute a diffusion scale space [1,9]. The isotropic diffusion scale space [10] is governed by the heat equation, which diffuses image isotropically in each local direction, so it smooths out details when removing noises. Perona and Malik [11] proposed anisotropic diffusion model with diffusion coefficients spatially varying to better preserve details. Rudin et al. [12] designed total variation diffusion scheme constrained by noise statistics [13,14]. Many other types of PDE diffusion filters were also designed [15–18].

In this paper, we concentrate on the anisotropic diffusion scale space, which is controlled by anisotropic diffusion equation [19]:

$$\frac{\mathrm{d}u(x,t)}{\mathrm{d}t} = \mathrm{div}[g(|\nabla u|)\nabla u]. \tag{1}$$

An essential problem of diffusion equation for denoising is how to select the optimal scale, i.e., stopping time, to get satisfactory result. This is usually called scale selection problem. It can be formulated as follows. Given the scale space $\{u(x,t)\}_{0 \le t < +\infty}$ induced by anisotropic diffusion equation, the aim is to find the optimal stopping time t^*, at which the denoised image $u(x,t^*)$ is most similar to the original non-degraded image. When selecting small stopping time, the noises are not adequately smoothed out, but when selecting large stopping time, image details are also removed. So the scale selection problem significantly affects the performance of diffusion filter, and there is a trade-off between removing noises and preserving details.

It is non-trivial to select the optimal scales for anisotropic diffusion filter. Much work have been done to solve this problem. Dolcetta and Ferretti [20] formulated the scale selection problem as the minimization of an energy function combining computing cost and stopping cost. Carefully balancing these two terms is necessary to achieve desired result. Weickert [21] proposed a signal to noise ratio (SNR) criterion to select the optimal stopping time t^*, at which the denoised image $u(x,t^*)$ satisfies

$$\frac{\mathrm{Var}(u(x,t^*))}{\mathrm{Var}(u(x,0))} = \frac{1}{1+1/\mathrm{SNR}}. \tag{2}$$

This criterion can well select the optimal scale because the SNR value of the noisy image is known. Solo [22] proposed to select scale by designing a simple quadratic measurement of the reconstruction quality, which can be computed by SURE (Stein's unbiased risk estimator). Mrazek [23] designed a scale selection criterion to choose the optimal scale when the correlation between the diffused image and removed noises is minimized:

$$t^* = \underset{\{t \ge 0\}}{\mathrm{argmin}} \frac{\mathrm{Cov}(u(x,0)-u(x,t),u(x,t))}{\sqrt{\mathrm{Var}(u(x,0)-u(x,t))\mathrm{Var}(u(x,t))}}. \tag{3}$$

* Corresponding author. Tel.: +86 29 82668005; fax: +86 29 82668559.
E-mail addresses: jiansun@mail.xjtu.edu.cn (J. Sun), zbxu@mail.xjtu.edu.cn (Z. Xu).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.02.019

This scheme is very simple and does not assume any noise statistics or SNR of the noisy image, however the correlation criterion is not strongly related to the denoising quality. Papandreou [24] developed a scale selection scheme by cross-validation. It divides the image region into training set and test set, the scale is selected by minimizing L_1 or L_2 error measurement on the test set. This scheme is well founded on the statistical learning theory, however, it is still hard to guarantee that the selected scale is optimal. Total variation (TV) model [12] uses noise statistics to constrain the TV diffusion process and makes it converge to the steady state, in which the removed noises satisfy known noise statistics. This can also be seen as a scale selection scheme, which is similar to SNR criterion.

Most of the previous work select a global stopping time for the entire image. The global scale selection scheme has the following problem. See Fig. 1 for example, the optimal scale is computed for the entire image region, a smooth region (in red rectangle) and a textured region (in blue rectangle). Scale is selected to minimize the mean squared error (MSE) between the diffused image and the original non-degraded image. Fig. 1(c) shows that the optimal scale of the smooth region is 42.2. The optimal scale of the textured region is 14.8, which is much smaller than that of the smooth region. This means that, the diffusion process should stop early in the textured region to better preserve image details and stop late in the smooth region to adequately remove noises. The optimal scale for the entire image is 25.0, which is the result of balancing the scale selection in both smooth regions and textured regions.

The above observation motivates us to design a segment-based scale selection model in this paper. We segment the image into small regions, and select different scales for different regions. The final diffused image is the composition of the segments at their optimal scales in the scale space, instead of a single anisotropic diffusion filtering result at a global optimal scale. The advantages of this scheme are as follows. Firstly, the textured regions and smooth regions are segmented and assigned labels separately. This overcomes the balancing effect over the textured and smooth regions caused by the global scale selection scheme. Secondly, this segment-based image representation helps us to design a better noise estimation algorithm, which is required by our scale selection model.

Considering scale coherence between neighboring segments, we propose a Markov random field (MRF) scale selection model finally. The MRF model is a popular framework for graph labeling problem in recent years, and was widely investigated in the image segmentation [25–28], digital photomontage [29], stereo vision [30], texture synthesis [31], and so on. Using this framework, we naturally incorporate scale coherence constraint over the neighboring segments into the scale selection problem. Then the scale selection model prefers the similar scale labels for appearance-similar neighboring segments and different scale labels for neighboring segments across image edges.

Compared with the previous scale selection schemes, our MRF scale selection model significantly improves the performance of anisotropic diffusion filter for image denoising. Experiments on natural images show that the image textures are better preserved and noises are adequately removed in the final denoised image. The peak signal to noise ratio (PSNR) values of our results by combining anisotropic diffusion with MRF scale selection model are comparable to or even better than the other state-of-the-art denoising algorithms.

The workflow of anisotropic diffusion filter combined with MRF scale selection model is illustrated in Fig. 2. Firstly we gradually diffuse the input image by anisotropic diffusion to construct a discrete scale space, and over-segment the diffused image at the largest scale T into segments with homogenous spatial coordinates and color/intensity values. Then we estimate the noise statistics by segment-based noise estimation algorithm. Using this statistics, we select scales for segments by MRF scale selection model. The final output image is the fusion of the diffusion scale space by composing the segments at their optimal scales in the scale space.

We will introduce each components of this workflow in the following sections. In Section 2, we briefly review the anisotropic diffusion filter and introduce how to construct anisotropic diffusion scale space. In Section 3, we firstly design the over-segmentation algorithm, then propose scale selection criteria for each segment, finally construct MRF model to fuse the diffusion scale space. In Section 4, we propose a segment-based noise estimation algorithm. In Section 5, comparisons with the pervious scale selection schemes and denoising algorithms are represented.

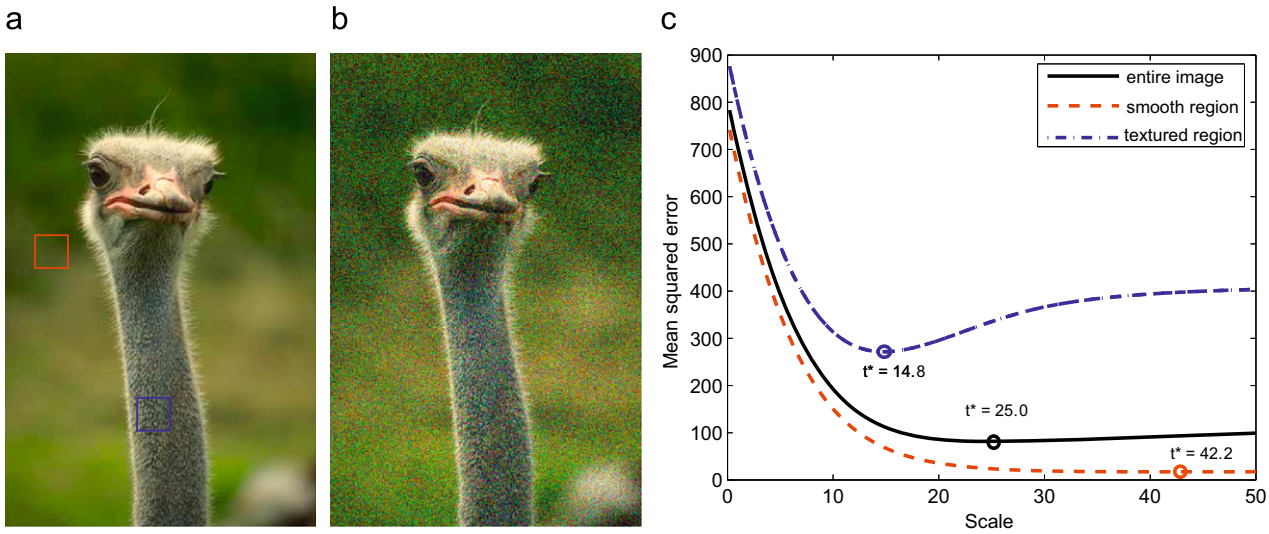

Fig. 1. Scale selection for image and its sub-regions: (a) original image; (b) noisy image with noise standard deviation 30 and (c) MSE with respect to increasing scales for entire image region, blue region and red region. The optimal scales for entire image region, red region and blue region are 25.0, 42.2 and 14.8. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

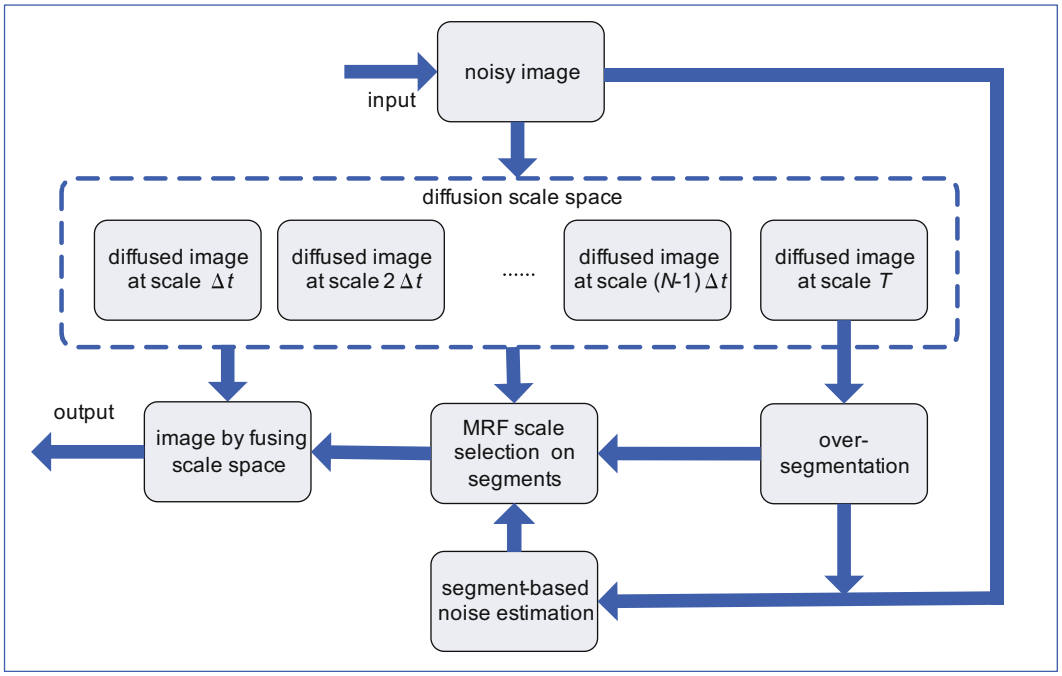

Fig. 2. Algorithm workflow. Noisy image is firstly diffused to generate a scale space, and over-segment the diffused image at largest scale T. Then we estimate the noise statistics by segment-based noise estimate algorithm. Using this statistics, we select scales for segments by MRF scale selection model. The final output image is the composition of the segments at their optimal scales in the scale space.

2. Anisotropic diffusion scale space

Anisotropic diffusion equation can be viewed as the gradient descent flow to minimize the energy functional:

$$E(u)=\rho(|\nabla u|^2). \tag{4}$$

Its gradient descent flow is controlled by anisotropic diffusion equation:

$$\frac{\mathrm{d}u(x,t)}{\mathrm{d}t}=\mathrm{div}[\rho'(|\nabla u|^2)\nabla u]. \tag{5}$$

It is same as defined by Eq. (1) when $g(|\nabla u|)=\rho'(|\nabla u|^2)$. The input noisy image is taken as its initial condition, and reflection boundary condition is used. $g(\cdot)$ is selected as a decreasing function which satisfies $g(r)\to 0$, as $r\to+\infty$.

As the scale t increases, the multi-scale analysis of the image is performed. The diffused images $\{u(x,t)\}_{0\le t<\infty}$ with increasing scales constitute a scale space, which is called anisotropic diffusion scale space. The steady state will be flat image or piecewise smooth image [32,33].

Numerically, this diffusion equation can be implemented by

$$\begin{aligned}u^{t+1}(i,j)=&\,u^t(i,j)+\tau[g(|\nabla u^t(i+1/2,j)|)\\&\times(u^t(i+1,j)-u^t(i,j))+g(|\nabla u^t(i-1/2,j)|)\\&\times(u^t(i-1,j)-u^t(i,j))+g(|\nabla u^t(i,j+1/2)|)\\&\times(u^t(i,j+1)-u^t(i,j))+g(|\nabla u^t(i,j-1/2)|)\times(u^t(i,j-1)-u^t(i,j))].\end{aligned} \tag{6}$$

We set $\tau=0.15$, and diffuse the image until scale T at which the image is over-smoothed (see Fig. 3(a) for example), then a discrete scale space can be constructed before this scale T. The scale T is selected when the variance of the removed noises is larger than the added noise variance (1.1 times of the added noise variance is set in our implementation). A basic wavelet domain noise estimation algorithm [34] is used to estimate the noise variance in this step, a more sophisticated noise estimation algorithm will be designed for scale selection in Section 4.

We further discretize the scale space into N frames, in which each frame is defined as a diffused image at a certain scale. So the scale interval is $\Delta t=T/N$. Then we construct a spatial and time discrete anisotropic diffusion scale space:

$$S=\{u(x,t)\}_{t=\Delta t,2\Delta t,\ldots,N\Delta t}. \tag{7}$$

3. Scale selection model

In this section, we introduce the segment-based scale selection model. Firstly we over-segment the image into a collection of segments with homogenous spatial coordinates and color/intensity values. Then we design the scale selection criteria to select scale for each segment. Finally, we develop a MRF scale selection model on the graph over image segments, and minimize it by tree-reweighted message passing algorithm.

3.1. Image over-segmentation

Directly over-segmenting the noisy image is not robust due to noises, so we over-segment the over-smoothed image at scale T into segments. We use k-means algorithm with five dimensional features consisted of color and spatial position (r,g,b,x,y) to over-segment the image [35,36]. To initialize the k-means algorithm, we regularly segment the image into 16×16 blocks as initial segmentation. Then iteratively update the clustering until convergence. Fig. 3(b) shows an example of over-segmentation.

Denote the image region as Ω, and each segment as S_i, $i=1,\ldots,K$, K is the total number of segments. Obviously, $\Omega=\bigcup_i S_i$. Given this representation, we will present our segment-based scale selection model.

3.2. Scale selection criteria for each segment

We now develop scale selection criteria to select scale for each segment. A basic observation is that the pixels in each image

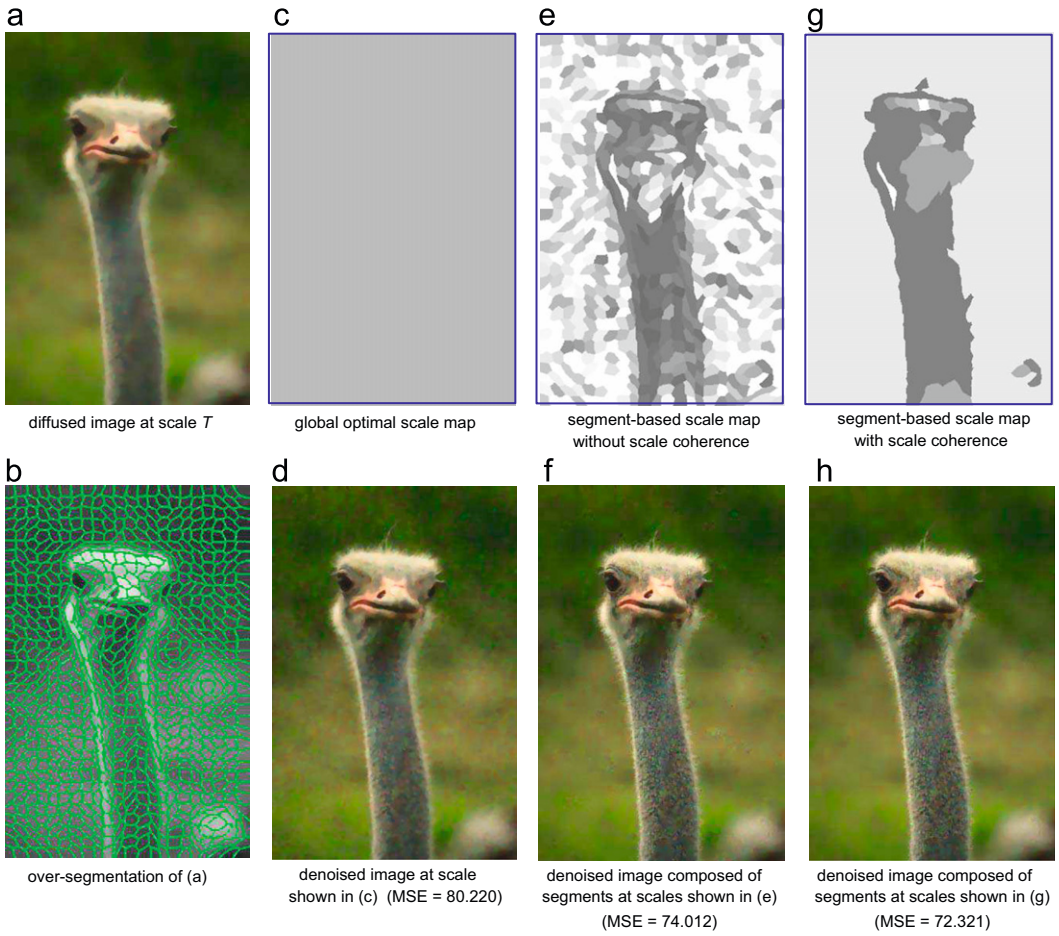

Fig. 3. Comparison between global scale selection scheme and segment-based scale selection model (the added noise standard deviation is 35): (a), (b) image over-segmentation at scale T, (c), (d) global optimal scale map and the denoised image, (e), (f) scale map and denoised image by segment-based scale selection without scale coherence, (g), (h) scale map and denoised image by MRF scale selection model.

segment are similar in their color/intensity values, so we can reasonably assume that image signals in each segment obey a Gaussian distribution.

We deal with additive noises in this paper. The image degradation by additive noises is modeled as

$$u_n = u^* + n, \tag{8}$$

where u^* is the original non-degraded image, u_n is the noisy image and n is assumed to be the additive Gaussian noise. In each segment S_i, we denote the variance and mean of u_n by Σ and μ, the noise variance of n by Σ_n (assumed to be known in this section), and the variance and mean of u^* by Σ_s and μ_s. Based on the image degradation model in Eq. (8), we deduce that

$$\Sigma = \Sigma_s + \Sigma_n, \tag{9}$$

$$\mu_s = \mu. \tag{10}$$

To select scale for image segment, we should measure the denoising quality of the diffused image at a certain scale. We first analyze the following error measurement between the diffused image at scale t and the original non-degraded image:

$$\mathrm{Err}(u^*(x),u(x,t)) = \mathrm{E}\{(u^*(x)-u(x,t))^2\}, \tag{11}$$

where $\mathrm{E}(\cdot)$ means expectation. The aim of scale selection is to minimize this error measurement. However the original non-degraded image is unknown, so we analyze the difference between the denoised image $u(x,t)$ and the noisy image $u_n(x)$ instead:

$$\begin{aligned}\mathrm{E}\{(u(x,t)-u_n(x))^2\} &= \mathrm{E}\{(u(x,t)-u^*(x)+u^*(x)-u_n(x))^2\}\\ &= \mathrm{E}\{(u(x,t)-u^*(x))^2\}+\mathrm{E}\{(u^*(x)-u_n(x))^2\}\\ &\quad +2\mathrm{E}\{(u(x,t)-u^*(x),u^*(x)-u_n(x))\}.\end{aligned} \tag{12}$$

In bias and variance trade-off analysis, $u^*(x)$ is assumed as deterministic constant, so the covariance term vanishes. But we have assumed that $u^*(x)$ is also a Gaussian distribution random variable in each segment. Therefore, to minimize $\mathrm{E}\{(u(x,t)-u^*(x))^2\}$, we expect

$$\begin{aligned}&[\mathrm{E}\{(u^*(x)-u_n(x))^2\}-\mathrm{E}\{(u(x,t)-u_n(x))^2\}]\\ &\quad +2\mathrm{E}\{(u(x,t)-u^*(x),u^*(x)-u_n(x))\}=0.\end{aligned} \tag{13}$$

We expect both of these two terms in Eq. (13) equal to zero:

$$\mathrm{E}\{(u_n(x)-u^*(x))^2\}-\mathrm{E}\{(u_n(x)-u(x,t))^2\}=0, \tag{14}$$

$$\mathrm{E}\{(u(x,t)-u^*(x),u^*(x)-u_n(x))\}=0, \tag{15}$$

where $\mathrm{E}\{(u_n(x)-u^*(x))^2\}=\Sigma_n$. The first term equals to zero when the removed noises $u_n(x)-u(x,t)$ obey the known noise distribution, i.e., $N(0,\Sigma_n)$. The second term equals to zero when $u(x,t)=u^*(x)$, which is unachievable because it is the target of image denoising. Instead of expecting $u(x,t)=u^*(x)$, we expect their distributions are same. We find that these two criteria are reasonable for scale selection and complementary to each other.

The first criterion means that the removed noises should obey the noise distribution, and the second criterion means that the signal of the final diffused image should obey the distribution of original image signal in each segment.

In summary, the criteria of scale selection for each segment S_i are

$$u_n(x)-u(x,t)\sim N(0,\Sigma_n), \tag{16}$$

$$u(x,t)\sim N(\mu_s,\Sigma_s), \tag{17}$$

in which $\mu_s=\mu$ and $\Sigma_s=\Sigma-\Sigma_n$. We use Kullback–Leibler (K–L) distance to model the distance between the estimated distribution and the target distribution. Take the second criterion as an example, we estimate the parameters of Gaussian distribution for each segment S_i as

$$\hat{\mu}_s=\frac{1}{|S_i|}\sum_{x\in S_i}u(x,t),$$

$$\hat{\Sigma}_s=\frac{1}{|S_i|}\sum_{x\in S_i}(u(x,t)-\hat{\mu}_s)(u(x,t)-\hat{\mu}_s),$$

where $|\cdot|$ means number of pixels. Then the K–L distance between the estimated image signal distribution and true image signal distribution for segment S_i at scale t_i is

$$\mathrm{KL}_s(t_i;S_i)=\sum_{x\in S_i}\left[G(u(x,t);\hat{\mu}_s,\hat{\Sigma}_s)\log_2\frac{G(u(x,t);\hat{\mu}_s,\hat{\Sigma}_s)}{G(u(x,t);\mu_s,\Sigma_s)}\right], \tag{18}$$

where $G(\cdot)$ is probability distribution function of Gaussian distribution. We can similarly define the distance $\mathrm{KL}_r(t_i;S_i)$ for the first criterion.

We combine these two K–L distances as the scale selection cost for segment S_i at scale t_i:

$$D_i(t_i;S_i)=\frac{1}{2}\frac{\mathrm{KL}_s(t_i;S_i)}{\sum_{t'}\mathrm{KL}_s(t';S_i)}+\frac{1}{2}\frac{\mathrm{KL}_r(t_i;S_i)}{\sum_{t'}\mathrm{KL}_r(t';S_i)}. \tag{19}$$

Then the optimal scale for S_i is selected as the scale minimizing $D_i(t_i;S_i)$:

$$t_i^*=\underset{t_i}{\mathrm{argmin}}\{D_i(t_i;S_i)\}. \tag{20}$$

After selecting optimal scales for all the segments, the final output diffused image is the composition of the segments at their optimal scales in the diffusion scale space.

Figs. 3(e) and (f) show the scale selection results by this scale selection criteria. Figs. 3(c) and (d) give the results of ideal case of global scale selection scheme, i.e., selecting scale when the diffused image is most similar to the original image. Compared with the ideal case of global scale selection scheme, the MSE value of the result by this criteria is much lower.

We find that selecting scale for each segment separately causes scale inconsistency between neighboring segments. See Fig. 3(c) for example, some appearance-similar and neighboring segments are assigned significantly different scales, which will cause artifacts in the final output diffusion result. That is because our scale selection model for each segment is to measure the similarity between the noises or signal distributions and the corresponding known distributions. But in some segments with typically hundreds of pixels, the statistical distributions cannot be estimated accurately due to limited samples. Then wrong scale labeling will appear.

3.3. Scale selection by super-pixel MRF model

We use Markov random field (MRF) [37,38] model to regularize the scale selection problem. Scale coherence between neighboring segments is considered in the MRF framework.

The MRF model is defined on a graph $G=(\mathcal{V},\mathcal{E})$. $\mathcal{V}=\{v_1,v_2,\ldots,v_K\}$ is denoted as a set of graph nodes. In this paper, the graph is constructed on image super-pixels, i.e., image segments. So each node v_i in the graph represents a segment S_i in the image, and the total number of segments is K. We focus our discussion on the pairwise Markov network, so $\mathcal{E}$ is a set of pairwise edges between these neighboring segments. This super-pixel MRF graph is illustrated in Fig. 4.

Each segment S_i is related to a random variable X_i, and the field of random variables is denoted as $X=\{X_1,X_2,\ldots,X_K\}$. Each random variable X_i takes a value t_i from the index set of frames $\{1,\ldots,N\}$ in the discrete scale space. Then $t=\{t_1,t_2,\ldots,t_K\}$ is a configuration of X, its joint probability is defined as

$$\Pr(X=t)\propto\prod_{i\in\mathcal{V}}\exp(-D_i(t_i;S_i))\prod_{(i,j)\in\mathcal{E}}\exp(-V_{i,j}(t_i,t_j;S_i,S_j)). \tag{21}$$

Maximizing this probability is equivalent to minimize its minus logarithm:

$$E(t)=\sum_{i\in\mathcal{V}}D_i(t_i;S_i)+\sum_{(i,j)\in\mathcal{E}}V_{i,j}(t_i,t_j;S_i,S_j). \tag{22}$$

The first term in Eq. (22) is same defined as Eq. (19). It is a penalty function measuring the cost of a segment S_i with scale label t_i. The second term in Eq. (22) is a pairwise smoothness function of two neighboring segments with scale label t_i and t_j, respectively.

For neighboring segments S_i and S_j, we define the smoothness function as

$$V_{i,j}(t_i,t_j;S_i,S_j)=[1-\delta(t_i,t_j)]\exp\left(-\frac{|\mu(S_i)-\mu(S_j)|^2}{2\langle(\mu(S_i)-\mu(S_j))\rangle^2}\right), \tag{23}$$

where

$$\delta(t_i,t_j)=\begin{cases}1 & \text{if } t_i=t_j,\\ 0 & \text{otherwise.}\end{cases} \tag{24}$$

$\mu(S_i)$ is the mean color or intensity on segment S_i, and $\langle\cdot\rangle$ means expectation. This term constrains the scale coherence between two appearance-similar segments S_i and S_j, and relaxes this tendency when S_i and S_j are significantly different. In other words, it penalizes different scale labeling between pairs of appearance-similar segments, and prefers different scale labeling for pairs of appearance-different segments. This type of definition of smoothness term is also investigated in [25–27,30].

The final MRF scale selection model is defined as

$$E(t)=\sum_i\left[\frac{1}{2}\frac{\mathrm{KL}_s(t_i;S_i)}{\sum_{t'}\mathrm{KL}_s(t';S_i)}+\frac{1}{2}\frac{\mathrm{KL}_r(t_i;S_i)}{\sum_{t'}\mathrm{KL}_r(t';S_i)}\right]+\lambda\sum_{j\in N(i)}\left[(1-\delta(t_i,t_j))\exp\left(-\frac{|\mu(S_i)-\mu(S_j)|^2}{\langle(\mu(S_i)-\mu(S_j))\rangle^2}\right)\right]. \tag{25}$$

After minimizing this energy function, the optimal scale for segment S_i is $t_i\times\Delta t$. Then the segment S_i in the output denoised image is coming from the same segment in the scale space frame at S_i's optimal scale.

The energy function of Eq. (25) is a typical model for graph-labeling problem in MRF framework, which is widely viewed as an intractable NP-hard problem. Recently, algorithms such as graph cut [39,40], loop belief propagation [41,42] or tree-reweighted message passing algorithm [43,44] have been designed to efficiently optimize this MRF energy function. It is proved that both graph-cut algorithm and tree-reweighted message passing algorithm (TRW) achieve high-quality optimization results [45] for MRF energy function.

In this paper, we use the TRW algorithm to optimize Eq. (25), which is a recently developed message passing algorithm for optimizing MRF energy function. This algorithm stems from the

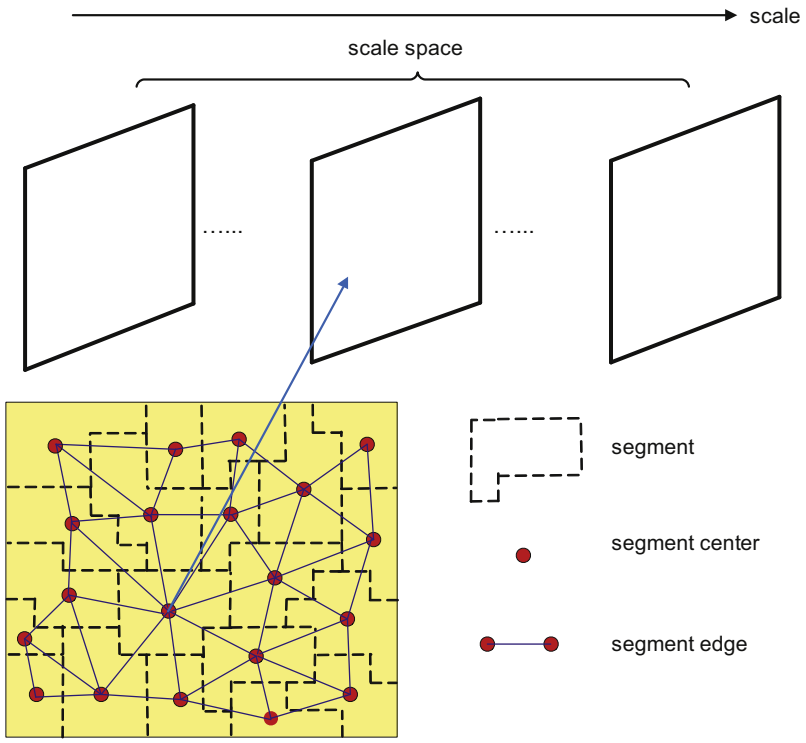

Fig. 4. Super-pixel Markov random field graph. The graph nodes are the segments of the image, and the graph edges are connections between neighboring segments. Each segment is assigned a scale in the scale space.

study of Wainwright et al. [43], in which they deduced a lower bound for MRF energy function, then tree-reweighted max-product message passing algorithm was designed to optimize the MRF energy by maximizing its lower bound. This algorithm consists of a class of message passing algorithms to find a collection of tree structured distributions with a common optimum. Kolmogorov [44] developed a modification of this algorithm, and it guarantees to increase the lower bound in each iteration. We use the improved version of TRW in our implementation, and the C++ code is available at http://www.adastral.ucl.ac.uk/~vladkolm/papers/TRW-S.html.

Figs. 3(g) and (h) show the scale selection results of the MRF scale selection model. Because scale coherence between neighboring segments is considered, a more reasonable scale map is inferred. It has smooth scale labels in the smooth regions and discontinuous scale labels across edges. Since the scale coherence is constrained and the segment boundaries are generally aligned with image edges, the final composition of the scale space has rare artifacts. The MSE of the final result with respect to the original image is further decreased compared with the result of segment-based scale selection without scale coherence.

4. Segment-based noise estimation

Same to the SNR scale selection criterion, our segment-based MRF scale selection model also assumes the noise variance is known. However in real cases, the noise variance is unknown. In this section, we will develop a segment-based noise estimation model in the wavelet domain.

We first review the methodologies of noise estimation. Some commercial softwares, for example NeatImage, estimate noise variance by selecting a smooth region from the image, then take the noise variance in that region as the noise variance of the image. Donoho [34] estimated noise variance by computing the median absolute deviation (MAD) value of the wavelet coefficients, which is a popular used method in image denoising. Stefano [46] proposed a training-based noise estimation method in the wavelet domain. Obviously, the accuracy of noise estimation relies on the selection of the training set. Ce [36] proposed a segment-based noise estimation model which focuses on CCD noises, i.e., intensity dependent noises. In this paper, we will develop a simple but effective noise estimation algorithm in the wavelet domain for additive noises.

Firstly, we introduce the idea of MAD noise estimation method. This method estimates the noise variance in the HH sub-band of the wavelet transform of the image, because the wavelet coefficients in that sub-band are mainly caused by the high-frequency noises. The squared root of noise variance is estimated as

$$\sigma = \mathrm{MAD}(W)/0.6745, \tag{26}$$

where W is the wavelet coefficients in the HH sub-band of wavelet transform. This method tends to over-estimate the noise variance. That is because the high-frequency details in the natural images will also produce non-zero wavelet coefficients in the HH sub-band of the wavelet transform. Especially when the noise variance is relatively smaller than the variance of the high-frequency details in the image, then the noise variance will be seriously over-estimated.

The over-estimation effect of the MAD noise estimation method can be alleviated by the segment-based image representation. The marginal distribution of the wavelet coefficients of natural image is generally a non-Gaussian heavy-tailed distribution [47–49]. This sparseness prior means that most of the wavelet coefficients of natural image are zero. So we discard the high-frequency regions of natural image and only use the low-frequency regions for better noise estimation.

This idea of noise estimation can be robustly implemented by the following procedures. We firstly estimate noise variance on each image segment by MAD noise estimation method. Discard the segments with variance largely deviating from the mean variance, then take the mean variance on the remaining segments

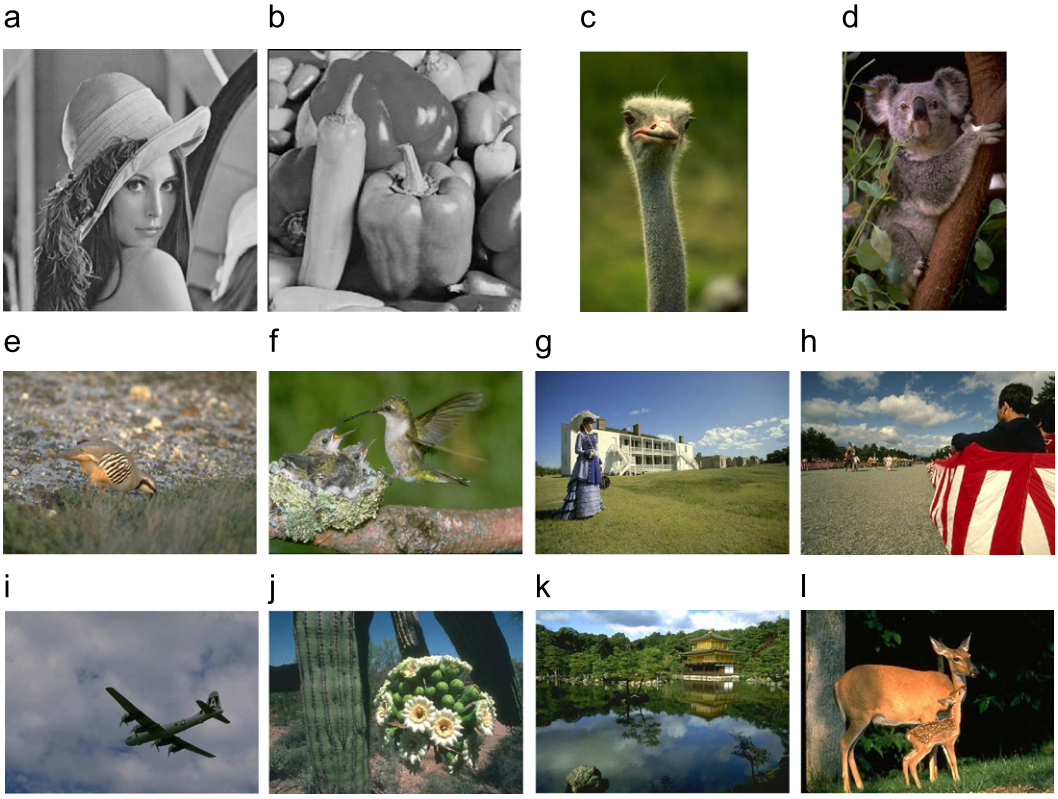

Fig. 5. Test images.

as the estimated noise variance. We iterate this procedure until convergence or reaching the maximal iterations (maximal 10 iterations is set in our implementation). The pseudo-code of this segment-based noise estimation algorithm is stated as follows.

Algorithm of segment-based noise estimation:

- *Input*: Noisy image u_n and its segments set $\{S_i\}$, $i=1,2,\ldots,K$.
- *Step* 1: Transform image by Daubechies wavelet transform. Denote the wavelet coefficients in the HH sub-band by W.
- *Step* 2: Estimate noise variance σ_i^2 by Eq. (26) on each segment S_i. Then compute the mean value μ_σ and standard deviation τ_σ on the set $\{\sigma_1,\sigma_2,\ldots,\sigma_K\}$. Initialize $\sigma=\mu_\sigma$, set $l=1$, $\varepsilon=0.01$, $e=2\varepsilon$, $L=10$.
- *Step* 3: While $l<L$ and $e>\varepsilon$, iterate the following procedures:
  Step 3a: Set $\sigma_{\text{old}}=\sigma$. Discard the segments with variance σ_i^2 satisfying $|\sigma_i-\mu_\sigma|>2\tau_\sigma$, then re-estimate μ_σ and τ_σ on the remaining segments.
  Step 3b: Set $\sigma=\mu_\sigma$, compute $e=|\sigma-\sigma_{\text{old}}|$, $l=l+1$.
- *Output*: Output the estimated noise variance σ^2.

5. Experiments

We use natural images to test the MRF scale selection model. These images are collected from the denoising benchmarks (http://decsai.ugr.es/~javier/denoise/index.html) and Berkeley segmentation data set (http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/). The Berkeley segmentation data set is composed of natural color images with hand-labeled segmentation masks. Recently, these natural images are also widely used for testing the algorithms in image denoising [50,51,36] and inpainting [50] due to their rich textures and colors. Fig. 5 lists all of the test images.

We will compare the performance of MRF scale selection model with the previous scale selection schemes, and also compare the denoising performance of anisotropic diffusion filter using MRF scale selection model with the state-of-the-art denoising algorithms. For the color images, the noise and signal distributions in R, G, B channels are assumed as independent, then Σ_s and Σ_n are diagonal three-dimensional matrices.

5.1. Noise estimation evaluation

We firstly evaluate the segment-based noise estimation algorithm. To measure the accuracy of noise estimation, we define the mean absolute deviation (MAD) between the estimated noise standard deviations and the true noise standard deviations as the error measurement. Under the added noise standard deviation σ_0, the MAD error of noise estimation is defined as

$$\text{MAD}(\sigma_0)=\frac{\sum_{i=1}^{12}|\sigma_i-\sigma_0|}{12}, \tag{27}$$

where σ_i is the estimated noise standard deviation for the i-th image in the test set.

We compare the segment-based noise estimation algorithm (Wav_Seg) with the classical wavelet domain MAD noise estimation algorithm (Wav). Table 1 lists the MAD errors for different noise levels[1] varying from 5 to 50. Fig. 6 plots these MAD errors. As we can see from Table 1 and Fig. 6, as the noise standard deviation increases, the segment-based noise estimation algorithm Wav_Seg is consistently more accurate than Wav algorithm. It significantly reduces the error of noise estimation especially when the noise standard deviation is low.

[1] Note that the noise level is the level of the noises statistics. It is defined as the standard deviation of noises in this paper.

Table 1
Noise estimation comparison.

Method	Noise standard deviation								
	5	10	15	20	25	30	35	40	50
Wav	1.3536	1.0447	0.8296	0.6721	0.5504	0.4501	0.3730	0.3113	0.2303
Wav_Seg	**0.8210**	**0.6500**	**0.5480**	**0.4476**	**0.3595**	**0.2835**	**0.2246**	**0.2140**	**0.2198**

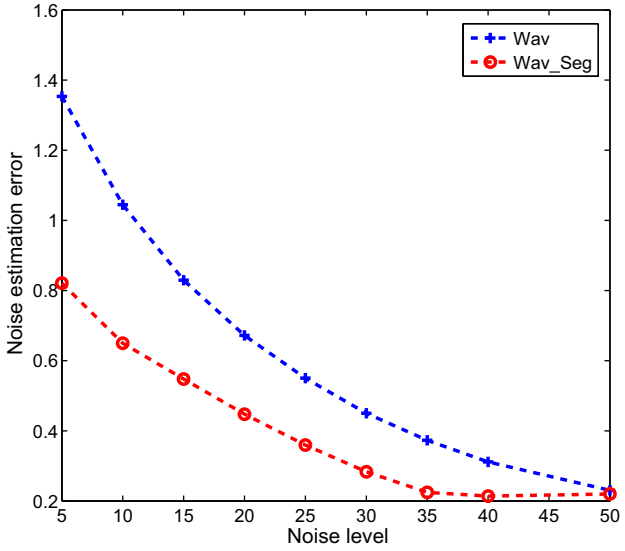

Fig. 6. Comparison between segment-based noise estimation in wavelet domain (Wav_Seg) and MAD noise estimation in wavelet domain (Wav). The noise estimation errors are plotted for different noise levels.

The question is, how does this segment-based noise estimation algorithm affect the performance of MRF scale selection model. It will be answered in the next subsection.

5.2. Scale selection comparison

In this subsection, we compare our MRF scale selection model with the previous global scale selection schemes. We will test two typical global scale selection schemes, i.e., SNR scheme and correlation scheme (Corr) for comparison. We will also present the results of the ideal case of global scale selection scheme (Opt), in which the optimal scale is selected when the diffused image is most similar to the original non-degraded image. Obviously, this is the limit of the global scale selection scheme, and all of the schemes in this category try to achieve this limit as close as possible. To test our proposed model, we will show the results of MRF scale selection model with the known noise variance (MRF1), and the estimated noise variance (MRF2) by the segment-based noise estimation algorithm, respectively.

In the experiments, we use two typical definitions of the diffusivity term $g(\cdot)$ in the anisotropic diffusion equation literatures [11,24],

$$\text{type 1}: \quad g(|\nabla u|) = \frac{1}{\sqrt{|\nabla u|^2+\varepsilon}}, \tag{28}$$

$$\text{type 2}: \quad g(|\nabla u|) = \frac{1}{1+|\nabla u|^2/Q^2}. \tag{29}$$

ε is set to make the denominator of the first type of diffusivity term non-zero, and it is set to be 0.01. The parameter Q in the second type of diffusivity term is set as in [11]. The anisotropic diffusion equations with these two types of diffusivity terms are equivalent to minimize the energy functionals in Eq. (4) with

$$\rho(|\nabla u|^2) = |\nabla u|, \tag{30}$$

$$\rho(|\nabla u|^2) = \log\left(1+\frac{|\nabla u|^2}{Q^2}\right), \tag{31}$$

respectively.

We evaluate the performance of a given scale selection scheme by the mean squared error (MSE) of the denoised image with scale (or scales) selected by this scale selection scheme. The MSE is defined as

$$\text{MSE}(u) = \frac{1}{|\Omega|}\sum_{p\in\Omega}(u(p)-u^*(p))^2, \tag{32}$$

where u is the denoised image, u^* is the original non-degraded image, and p is the pixel located in the image domain Ω. Lower MSE means better scale selection performance. We also design MSE ratio R to measure the relative performance between a given scale selection scheme and the ideal case of global scale selection. R is defined as

$$R = \frac{\text{MSE}_{\text{den}}}{\text{MSE}_{\text{opt}}}, \tag{33}$$

where MSE_{den} is the MSE of the denoised image obtained by the given scale selection scheme, and MSE_{opt} is the MSE of the denoised image in the ideal case of global scale selection scheme. $R<1$ means that the scale selection scheme is better than the ideal case of global scale selection scheme. Obviously, for all the global scale selection schemes, $R\geq 1$.

Tables 2 and 3 show the mean MSE values of the denoised images with respect to increasing noise levels varying from 5 to 50. Under each noise level, the mean MSE is the average MSE value of the denoising results of the test images. Tables 2 and 3 list the mean MSE values for anisotropic diffusion equation with types 1 and 2 diffusivity terms. In each table, we compare the performances of the five types of scale selection schemes, including SNR, Corr, Opt, MRF1, and MRF2. Fig. 7 is the plot of the MSE values. The left and right sides of Fig. 7 are performance comparison for the anisotropic diffusion equation with types 1 and 2 diffusivity terms, respectively. Tables 2, 3 and Fig. 7 tell us that, firstly the mean MSE values of SNR and Corr scale selection schemes are higher than the mean MSE values in the ideal case of global scale selection scheme. Secondly, the MRF scale selection models (MRF1 and MRF2) achieve much lower MSE values compared with the SNR and Corr based scale selection schemes. Thirdly, both MRF1 and MRF2 are better than the ideal case of global scale selection scheme in performance.

Tables 4 and 5 give the MSE ratios of the denoised images with respect to different noise levels for diffusion equation with types 1 and 2 diffusivity terms, respectively. The scale selection of the denoised images is performed by SNR, Corr, MRF1 and MRF2 scale selection schemes. The last column of each table lists the mean MSE ratios across different noise levels for these scale selection schemes. The left and right sides of Fig. 8 are the plots of these MSE ratios for diffusion equation with diffusivity types 1 and 2,

Table 2
Mean MSE for type 1 diffusivity term.

Method	Noise level								
	5	10	15	20	25	30	35	40	50
SNR	16.258	44.144	73.950	103.556	131.045	157.090	182.585	206.726	250.402
Corr	55.094	64.255	79.535	124.524	139.796	154.457	168.390	181.588	220.706
Opt	**12.412**	32.656	53.715	73.988	93.084	110.977	127.791	143.627	172.796
MRF1	12.612	**31.131**	**50.004**	**68.577**	86.741	103.945	120.833	**137.108**	**169.261**
MRF2	13.396	31.719	50.504	68.741	**86.369**	**103.627**	**112.421**	137.394	169.930

Table 3
Mean MSE for type 2 diffusivity term.

Method	Noise level								
	5	10	15	20	25	30	35	40	50
SNR	13.999	37.491	70.640	106.106	137.615	172.433	208.219	244.051	317.195
Corr	16.457	42.785	71.849	117.948	141.083	166.934	187.550	201.046	293.102
Opt	12.323	34.521	59.715	87.827	111.811	137.992	164.459	191.419	245.767
MRF1	**11.909**	**32.511**	**54.834**	78.844	98.945	120.530	141.695	162.591	**206.032**
MRF2	12.138	32.653	54.842	**76.836**	**98.403**	**119.639**	**140.629**	**161.689**	208.264

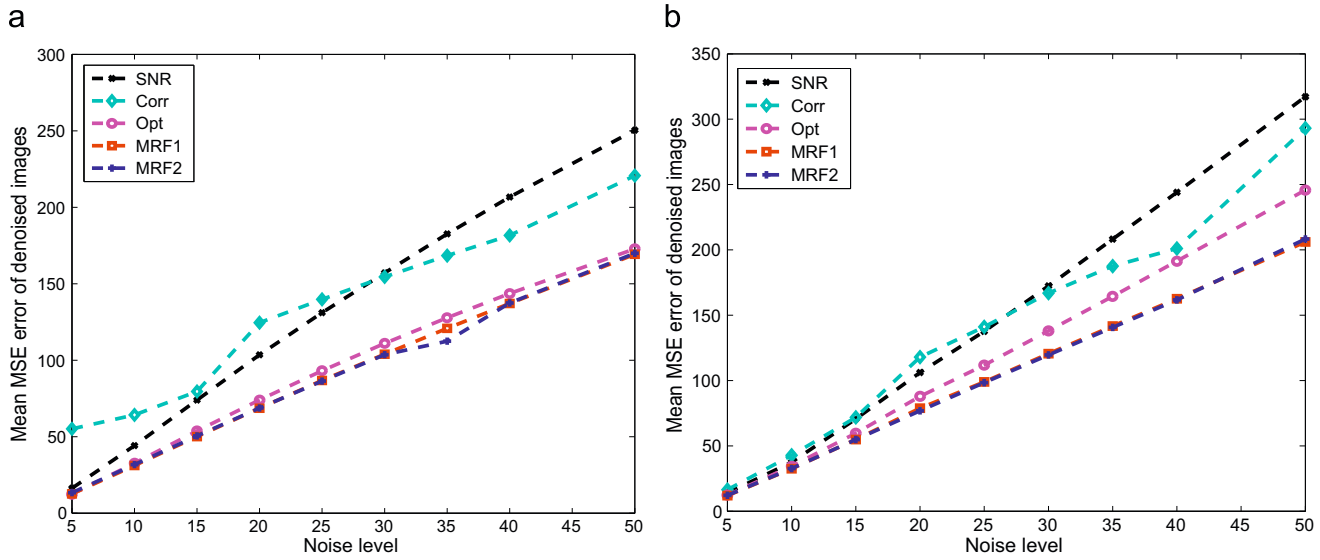

Fig. 7. Mean MSE of the denoised images by anisotropic diffusion filter using different scale selection schemes. The left and right sub-figures are for the diffusion equations with types 1 and 2 diffusivity terms, respectively.

Table 4
Mean MSE ratio for type 1 diffusivity term.

Method	Noise level									Mean
	5	10	15	20	25	30	35	40	50	
SNR	1.366	1.421	1.441	1.466	1.466	1.469	1.476	1.482	1.483	1.452
Corr	3.979	1.793	1.391	1.669	1.486	1.375	1.300	1.246	1.276	1.724
MRF1	**1.002**	**0.945**	**0.928**	**0.926**	0.933	0.941	0.952	**0.962**	**0.994**	**0.954**
MRF2	1.053	0.959	0.935	0.930	**0.933**	**0.940**	**0.900**	0.972	1.003	0.958

respectively. From Tables 4 and 5, we observe that the mean MSE ratios of SNR scheme and Corr scheme are 1.426 and 1.864 for type 1 diffusivity term, 1.249 and 1.226 for type 2 diffusivity term. They are all larger than 1 because these two schemes are definitely worse than the ideal case of global scale selection scheme. However, the mean MSE ratios of MRF1 and MRF2 are all lower than 1 for both two types of diffusivity terms. It means that MRF scale selection schemes (MRF1 and MRF2) are better than the ideal case of global scale selection scheme in mean performance.

We also find that the performance of MRF2, which estimates noise variance by the segment-based noise estimation algorithm, is nearly same to the performance of MRF1, which is given by the

Table 5
Mean MSE ratio for type 2 diffusion term.

Method	Noise level									Mean
	5	10	15	20	25	30	35	40	50	
SNR	1.166	1.103	1.233	1.259	1.281	1.298	1.312	1.318	1.330	1.256
Corr	1.270	1.196	1.169	1.333	1.238	1.183	1.120	1.043	1.171	1.191
MRF1	**0.967**	**0.945**	**0.919**	0.898	0.882	0.871	0.860	**0.846**	**0.837**	**0.892**
MRF2	0.982	0.947	0.920	**0.883**	**0.882**	**0.866**	**0.857**	0.847	0.851	0.893

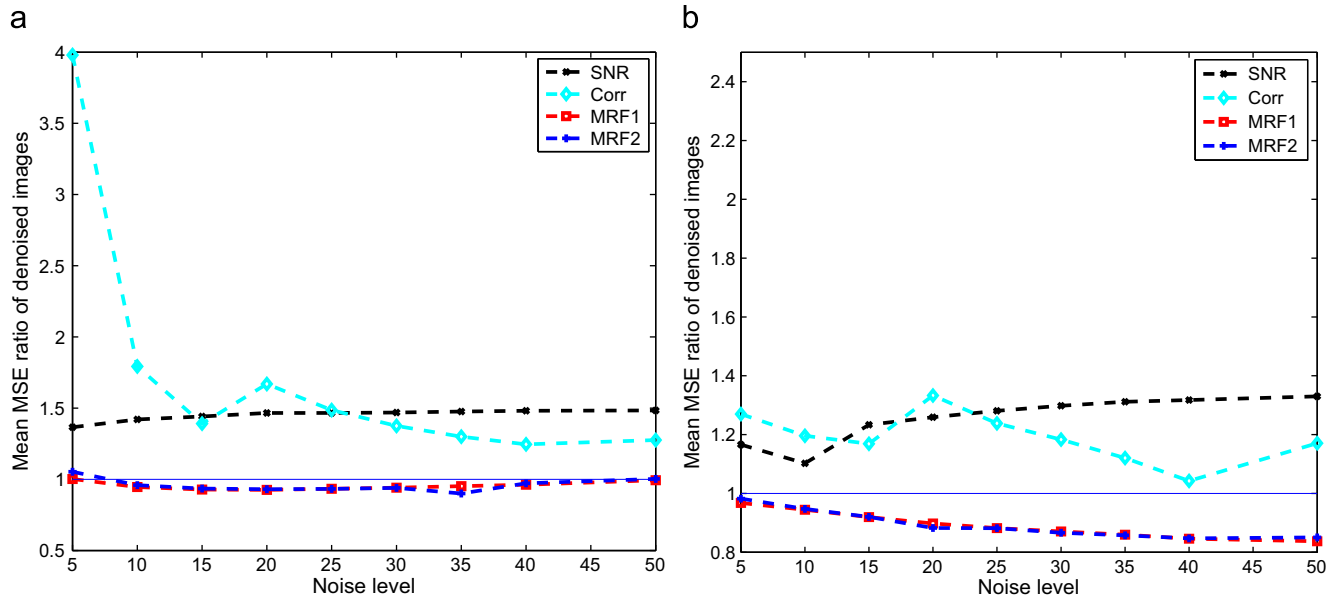

Fig. 8. Mean MSE ratios of the denoised images by anisotropic diffusion filter using different scale selection schemes. The left and right sub-figures are for the diffusion equations with types 1 and 2 diffusivity terms, respectively.

known noise variance. See Tables 4 and 5, the mean MSE ratios over all the test images and noise levels for MRF1 and MRF2 is 0.948 and 0.954 for type 1 diffusivity term, 0.895 and 0.898 for type 2 diffusivity term. It means that, the performance deficiency is very limited using the estimated noise variance by segment-based noise estimation algorithm. The MRF scale selection model with the estimated noise variance is still better than the ideal case of global scale selection scheme in performance.

5.3. Parameter setting

We now discuss the parameter settings for the MRF scale selection model. The first parameter is the number of frames N in the scale space, and the second parameter is the regularization parameter λ in Eq. (25).

5.3.1. Number of frames

Firstly, we discuss the setting of number of frames N in the scale space. The setting of N is equivalent to the setting of the scale interval Δt between consecutive frames due to their relationship $\Delta t = T/N$, in which T has been discussed in Section 2.

We set N from the noisy/original image pairs over the test set. For a given pair of noisy/original images, the scale interval Δt should be able to distinguish the different scales of the segments appeared in the noisy image. Here, the optimal scale of a segment is defined as the optimal stopping time when the diffused segment is most similar to the original non-degraded segment in MSE measurement. Inspired by this idea, a histogram of scale differences over pairs of segments is computed in the following set:

$$\{|t_{gt}(S_i) - t_{gt}(S_j)|; i,j \in [1,K], i \neq j, \text{ and } t_{gt}(S_i) \neq t_{gt}(S_j)\} \quad (34)$$

where $t_{gt}(S_i)$ is the optimal scale of the segment S_i. Then we set Δt as the value of bin where the integral of the histogram is 0.05. The number of frames is further set to $N = \lceil T/\Delta t \rceil$. In this way, 95% of the scale differences between segments in the noisy image can be distinguished.

We have estimated the mean scale interval and mean number of frames on the test set under different noise levels. These data are listed in Table 6. For both types 1 and 2 diffusivity terms, the mean scale interval increases with respect to noise levels. However, the mean number of frames is relatively stable with respect to the noise level, because the maximal scale T of the scale space also increases for larger noise level. So we set N to be the mean number of frames across different images and different noise levels, that is, 35 for the type 1 diffusivity term, and 41 for the type 2 diffusivity term.

We also investigate the influence of N on the scale selection performance. This is shown in Fig. 9, it plots the mean MSE over the test set with increasing N. We use the diffusion equation with the type 1 diffusivity term and the added noise level is 30. When N increases from 5 to 100, the mean MSE of the denoised images is not sensitive to N, and varies in a small interval [111.34,112.78]. Moreover, larger N does not always mean higher performance. That is because, larger N or smaller Δt may cause the scale selection problem to be more sensitive to the errors of scale labeling introduced by the energy term defined on individual

Table 6
Mean scale interval and mean number of frames for different noise levels.

Diffusivity type	Noise level								
	5	10	15	20	25	30	35	40	50
Type 1									
Mean Δt	0.140	0.267	0.467	0.547	0.760	0.940	1.120	1.307	1.500
Mean N	35	38	34	39	35	35	35	34	38
Type 2									
Mean Δt	0.072	0.126	0.176	0.216	0.288	0.324	0.376	0.400	0.410
Mean N	42	47	45	42	38	38	38	38	44

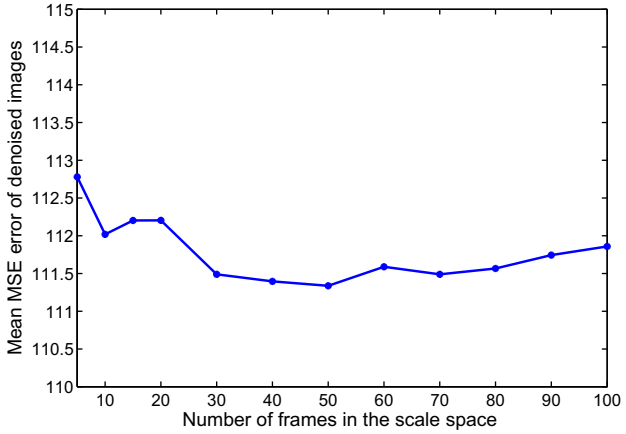

Fig. 9. Influence of number of frames on denoising performance. The added noise standard deviation is 30. The diffusion equation with the type 1 diffusivity term is used.

segments, i.e., the first term of Eq. (25). This is similar to the over-fitting phenomena in the problem of data classification.

5.3.2. Regularization parameter λ

Next we discuss how to set the regularization parameter λ in Eq. (25). This parameter balances the scale selection costs on individual segments and the scale coherence constraints between neighboring segments. When setting smaller λ, it will produce scale inconsistency between neighboring segments, and artifacts may be introduced in the final scale space fused image. Setting larger λ will pose tighter constraints on the scale coherence, which makes the scale labeling more reasonable. However it may assign the same scale label to the segments at different scales. The question now is how to design a criterion for selecting λ to better balance these two effects.

We select the parameter λ with respect to noise level σ using the following formula:

$$\lambda^*(\sigma) = \min_{\lambda} \left\{ \frac{1}{12} \sum_{i=1}^{12} \mathrm{MSE}(u_i(x;\lambda,\sigma)) \right\}, \tag{35}$$

where $u_i(x;\lambda,\sigma)$ is the denoised image for the i-th test image, given the regularization parameter λ and added noise standard deviation σ. The optimal λ^* is selected when the mean MSE between the denoised images and the non-degraded images in the test set is minimized. This is a commonly used discriminative parameter setting method. Table 7 lists the optimal regularization parameters $\lambda^*(\sigma_l)$ for discrete noise levels σ_l varying from 2 to 70, for diffusion equation with types 1 and 2 diffusivity terms, respectively. It is shown that λ^* increases with respect to noise levels. Then for given noise level σ_0, the $\lambda^*(\sigma_0)$ can be estimated by second order polynomial interpolation (if $\sigma_0 \in [2,70]$) or second order polynomial extrapolation (if $\sigma_0 \notin [2,70]$) among these data in Table 7.

Table 7
Optimal λ ($\times 10^{-3}$) under different noise levels for types 1 and 2 diffusivity terms.

Noise level	2	5	10	15	20	25	30	35	40	50	60	70
λ (type 1)	0.01	0.05	0.09	0.30	1.4	1.6	2.0	2.7	4.4	6.4	8.3	9.6
λ (type 2)	0.12	0.3	0.8	1.4	1.6	2.2	2.4	2.7	3.0	3.6	4.2	5.1

5.4. Denoising comparison

Now we compare the denoising performance of the anisotropic diffusion filter equipped with the MRF scale selection model to the other state-of-the-art denoising methods.

These denoising methods are introduced as follows. Total variation (TV) denoising model [12–14] is a classical and widely investigated image filter. It minimizes the image total variation while constraining the statistics of the removed noises. The non-local (NL) denoising algorithm [52] is a recently proposed image filtering algorithm. It performs filtering for a pixel by the expected color/gray value over the patches positioned in the non-local image domain, and the expectation is weighted by their similarities to the patch surrounding the pixel. In the experiments, we set the size of non-local image region for searching the patches as 21×21, the size of patch as 7×7, and the filtering parameter for computing expectation weights is set to the 10 times of noise standard deviation. The wavelet domain Gaussian scale mixture (GSM) denoising algorithm [49] uses Gaussian scale mixture model to fit the heavy-detailed distributions of the wavelet coefficients. Then the restored coefficients are inferred by Bayesian least square. We use steerable pyramidal decomposition with five scales to get the denoising results as in [49]. The wavelet domain GSM denoising algorithm has been reported to achieve the state-of-the-art performance in image denoising.

Figs. 10–12 present three examples of the denoising results by these algorithms (see the electrical version and zoom in to compare the results). Total variation method over-smooths the textured regions, for example the body of ostrich in Fig. 10, the trees in Fig. 11, and the grass in Fig. 12, but the noises in smooth regions are not adequately smoothed out. The result of non-local denoising algorithm has some artifacts both in the textured and smooth regions. The wavelet domain GSM algorithm can better balance the tasks of preserving textures and removing noises, but it causes significant ringing artifacts in the restored image. For the anisotropic diffusion filter combined with MRF scale selection model (MRF), we also present the image over-segmentation results and the scale maps on the segments

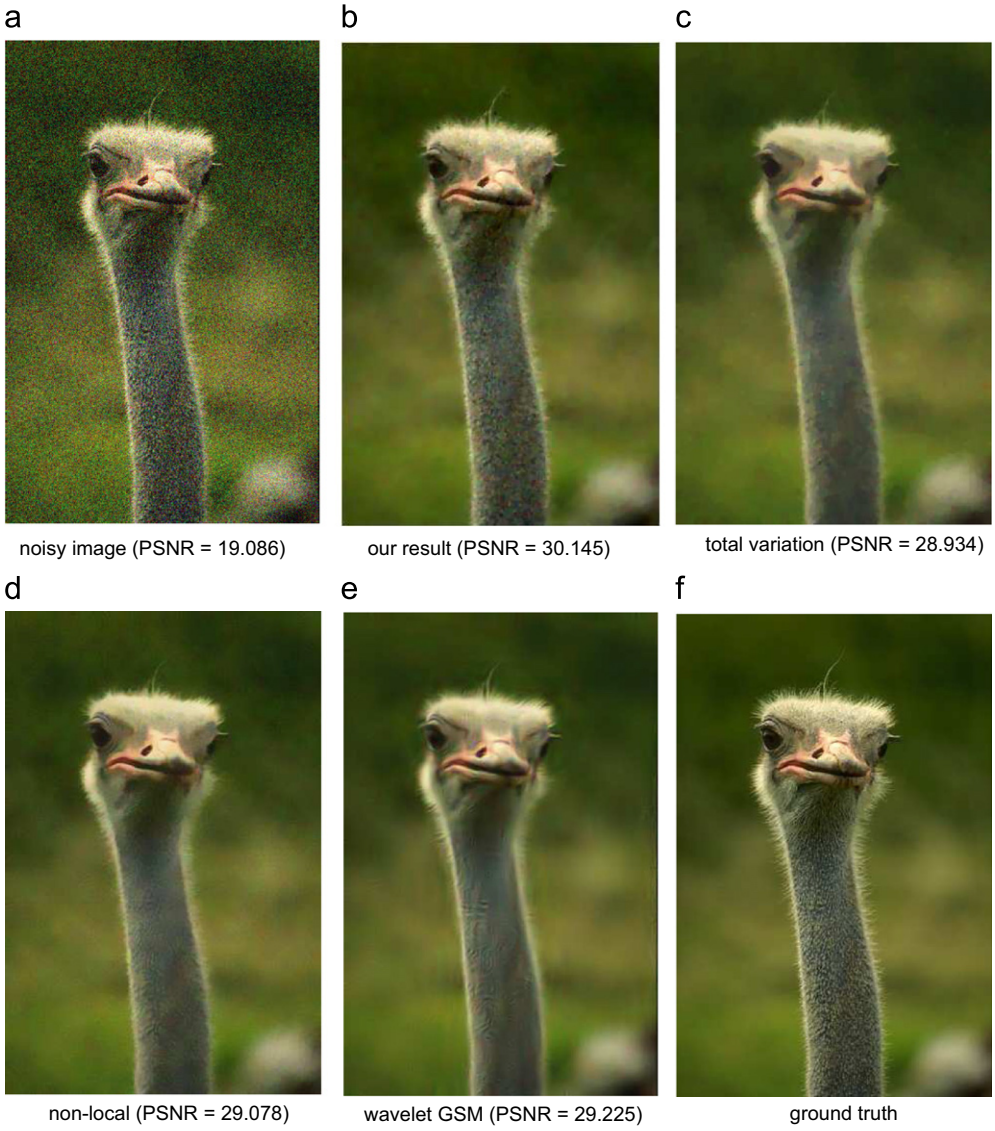

Fig. 10. Image denoising results comparison (zoom in for better comparison). The PSNR value of input noisy image is 19.086.

(lighter gray-scale means larger scale). The final denoising results show that our method can remove noises adequately in the smooth regions and preserve the textures in textured regions at the same time.

We also quantitatively measure the denoising performance of these algorithms by peak signal-to-noise ratio (PSNR) on the test images. Table 8 lists the PSNR values of results by TV, NL, GSM and the anisotropic diffusion filter (with type 1 diffusivity term) combined with MRF scale selection model (MRF). The last row of Table 8 shows the mean PSNR values by these denoising methods across different images. As we can see, both anisotropic diffusion filter combined with the MRF scale selection model and wavelet domain GSM achieve the best results. When the input image PSNR is 22.11 and 19.08, our model achieves six best results among 12 natural images, and achieves the best mean PSNR values. TV denoising model is a type of anisotropic diffusion filter with the same defined diffusivity term as the type 1 diffusivity term (Eq. (28)). Compared with the results of TV filter, our MRF scale selection model significantly increases the mean performance of anisotropic diffusion filter by 0.911 and 0.441 in mean PSNR, when the PSNR of noisy images are 22.110 and 19.086, respectively.

6. Conclusion

This paper presents a segment-based scale selection model for anisotropic diffusion scale space, in which the final denoised image is the fusion of the scale space. Scale coherence between neighboring segments is also considered, then a MRF scale selection model is designed finally, which is minimized by tree-reweighted message passing algorithm. In this model, textured region tends to be assigned smaller stopping time, and smooth region tends to be assigned larger stopping time, so the details are better preserved while noises are adequately smoothed out. We experiment our model on natural images, results show that MRF scale selection model is much better than the previous global scale selection schemes, and also better than the ideal case of the global scale selection scheme. Moreover, using MRF scale selection model, the denoising performance of anisotropic diffusion filter is significantly improved, and is comparable to or even better than the state-of-the-art denoising algorithms, e.g., non-local denoising method, total variation method and wavelet domain Gaussian scale mixture method.

Though segments can well distinguish image smooth regions or textured regions, some tiny details (for example the long hairs)

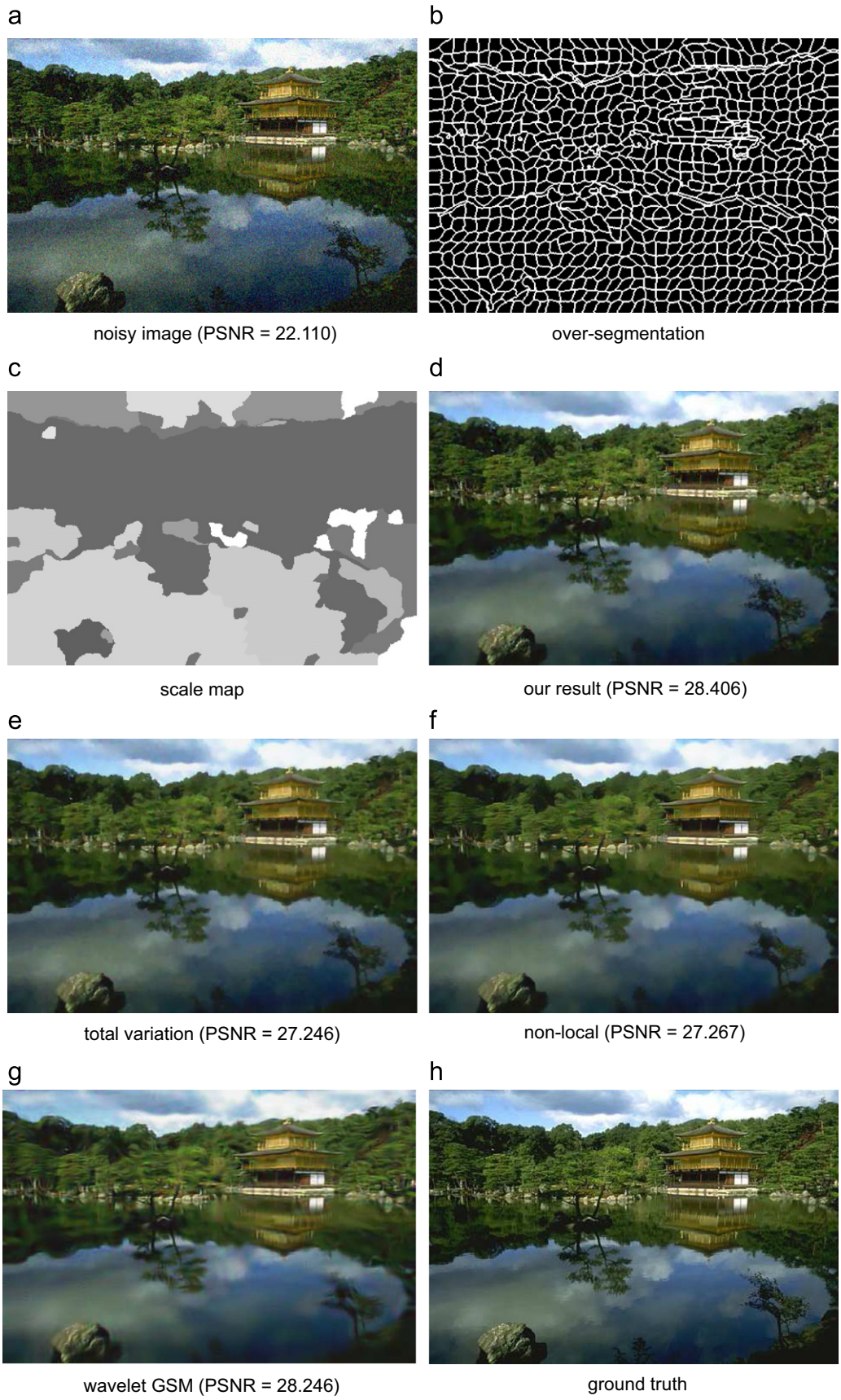

Fig. 11. Image denoising results comparison (zoom in for better comparison). The PSNR value of input noisy image is 22.110.

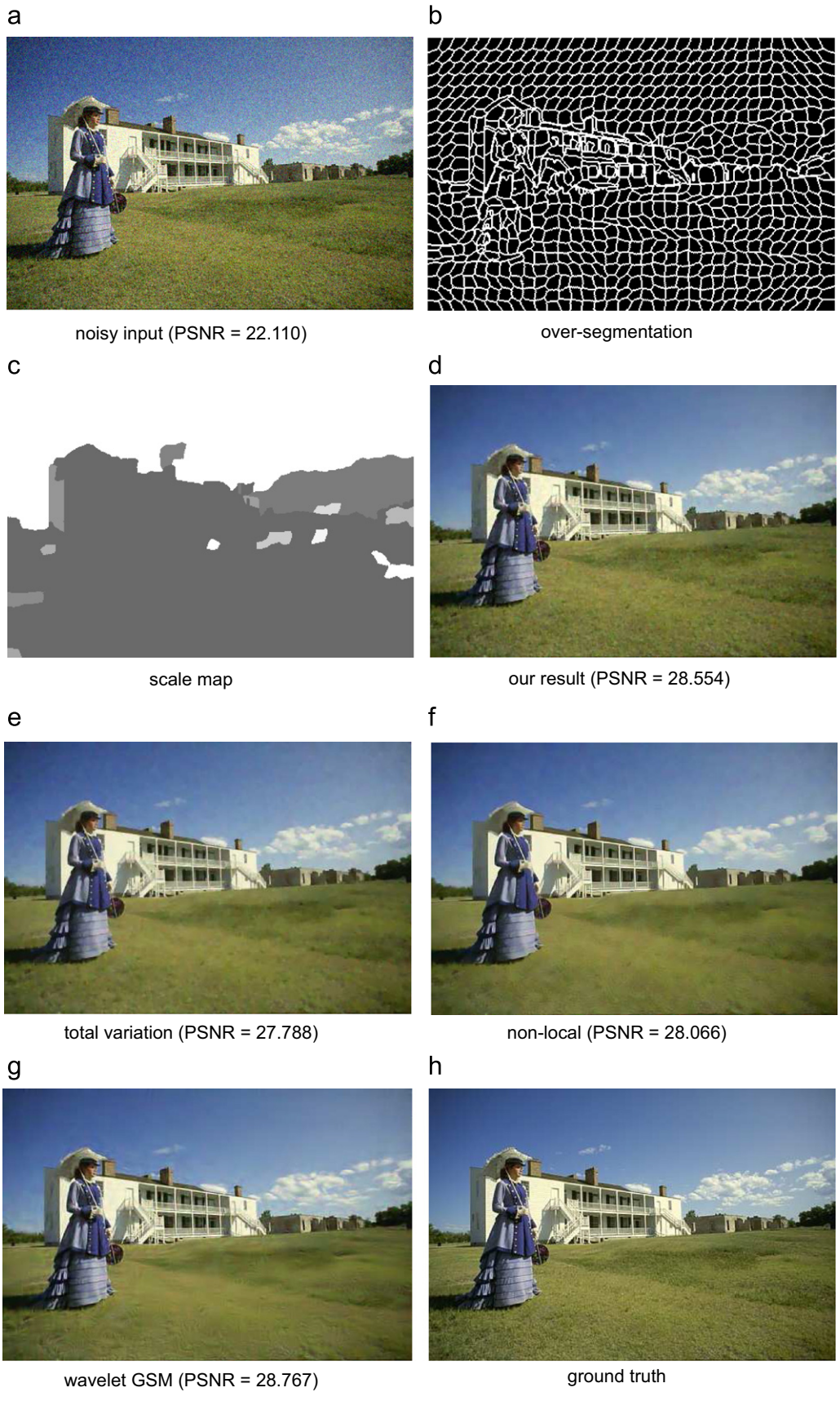

Fig. 12. Image denoising results comparison (zoom in for better comparison). The PSNR value of input noisy image is 22.110.

Table 8
PSNR by different denoising methods.

Image	PSNR=22.110				PSNR=19.086			
	TV	NL	GSM	MRF	TV	NL	GSM	MRF
(a)	30.400	31.358	**32.667**	31.650	29.473	29.239	**30.910**	29.932
(b)	28.853	29.883	**30.305**	30.084	27.644	27.666	**28.220**	27.981
(c)	30.862	30.888	31.113	**31.730**	28.934	29.078	29.225	**30.145**
(d)	28.840	29.218	29.441	**29.873**	27.806	27.061	27.204	**28.153**
(e)	29.163	29.665	**31.567**	30.788	28.187	27.637	**29.493**	28.934
(f)	27.898	27.838	**29.266**	28.920	26.406	26.065	27.072	**27.080**
(g)	27.788	28.066	**28.767**	28.554	26.553	26.525	**27.040**	26.853
(h)	27.699	28.035	27.997	**28.167**	26.235	26.167	**25.731**	26.202
(i)	36.917	35.549	36.984	**37.505**	35.137	32.756	34.728	**35.393**
(j)	27.652	27.405	**28.359**	28.328	26.214	25.848	26.436	**26.562**
(k)	27.246	27.267	28.246	**28.406**	26.067	25.606	25.839	**26.549**
(l)	28.529	28.314	28.518	**29.241**	27.067	26.464	**27.700**	27.459
Mean	29.360	29.456	30.269	**30.271**	28.052	27.509	28.300	**28.436**

are still hard to be represented. This makes our scale selection model not fine enough for some small details, and affects the final denoising performance. So it is necessary to investigate more accurate image super-pixel representation in the future. We also believe that, this model can be applied to other parameter dependent denoising algorithms, in which parameter can be seen as the scale. So applying this framework to other scale dependent algorithms is also our research direction.

Acknowledgments

The authors would like to thank the reviewers for their constructive comments and useful suggestions on this paper. This work was supported by the National Key Fundamental Research Program (the 973 Program) of China under Grant 2007CB311002, and the Key Project of the NSF of China under Grant 70531030.

References

[1] A.P. Witkin, Scale space filtering, in: Proceedings of International Joint Conference on Artificial Intelligence, 1983, pp. 1019–1021.
[2] T. Lindeberg, Edge detection and ridge detection with automatic scale selection, International Journal of Computer Vision 30 (2) (1998) 117–154.
[3] J.H. Elder, S.W. Zucker, Local scale control for edge detection and blur estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (7) (1998) 699–716.
[4] L.M. Kennedy, M. Basut, Scale space contours and localization property of a Gaussian derivative edge enhancement operator, in: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Orlando, FL, USA, 1997, pp. 643–648.
[5] T. Lindeberg, Feature detection with automatic scale selection, International Journal of Computer Vision 30 (2) (1998) 79–116.
[6] H.D. Cheng, J. Li, Fuzzy homogeneity and scale-space approach to color image segmentation, Pattern Recognition 36 (7) (2003) 1545–1562.
[7] R. Megret, J.M. Jolion, Tracking scale-space blobs for video description, IEEE Multimedia 9 (2) (2002) 34–43.
[8] Y. Leung, J.-S. Zhang, Z.-B. Xu, Clustering by scale-space filtering, IEEE Transactions on Pattern Recognition and Machine Intelligence 20 (12) (2000) 1396–1410.
[9] T. Lindeberg, Scale-space Theory in Computer Vision, Kluwer Academic Publishers, Netherlands, 1994.
[10] J. Babaud, A.P. Witkin, M. Baudin, R.O. Duda, Uniqueness of the Gaussian kernel for scale-space filtering, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1) (1986) 26–33.
[11] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (7) (1990) 629–639.
[12] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259–268.
[13] T.F. Chan, S. Osher, J. Shen, The digital TV filter and nonlinear denoising, IEEE Transactions on Image Processing 10 (2) (2001) 231–241.
[14] T.F. Chan, S.H. Kang, J. Shen, Total variation denoising and enhancement of color images based on the CB and HSV color models, Journal of Visual Communication and Image Representation 12 (4) (2001) 422–435.
[15] M. Black, G. Sapiro, D.H. Marimont, D. Heeger, Robust anisotropic diffusion, IEEE Transactions on Image Processing 7 (3) (1998) 421–432.
[16] G. Sapiro, Geometric Partial Differential Equations and Image Analysis, Cambridge University Press, Cambridge, 2001.
[17] D. Tschumperle, Fast anisotropic smoothing of multi-valued images using curvature-preserving PDE's, International Journal on Computer Vision 68 (1) (2006) 65–82.
[18] G. Gilboa, N. Sochen, Y.Y. Zeevi, Variational denoising of partly-textured images by spatially varing constrains, IEEE Transactions on Image Processing 15 (8) (2006) 2281–2289.
[19] J. Weikert, Anisotropic Diffusion in Image Processing, Teubner-Verlag, Stuttgart, Germany, 1998.
[20] I.C. Dolcetta, R. Ferretti, Optimal stopping time formulation of adaptive image filtering, Applied Mathematics and Optimization 43 (2001) 245–258.
[21] J. Weickert, Coherence-enhancing diffusion of colour images, Image and Vision Computing 17 (3) (1999) 201–212.
[22] V. Solo, A fast automatic stopping criterion for anisotropic diffusion, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal, Orlando, FL, USA, May 13–15, 2002, pp. 1661–1664.
[23] P. Mrazek, Selection of optimal stopping time for nonlinear diffusion filtering, International Journal of Computer Vision 52 (2/3) (2003) 189–203.
[24] G. Papandreou, P. Maragos, A cross-validatory statistical approach to scale selection for image denoising by nonlinear diffusion, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, June 20–25, 2005, pp. 625–630.
[25] A. Blake, C. Rother, M. Brown, P. Perez, P. Torr, Interactive image segmentation using an adaptive GMMRF model, in: Proceedings of European Conference on Computer Vision (ECCV), Prague, Czech, 2004, pp. 428–441.
[26] Y. Boykov, M.P. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images, in: Proceedings of International Conference on Computer Vision (ICCV), vol. 1, Vancouver, BC, Canada, 2001, pp. 105–112.
[27] Y. Li, J. Sun, C.K. Tang, H.Y. Shum, Lazy snapping, ACM Transaction on Graphics 23 (3) (2004) 303–308.
[28] J. Sun, J. Sun, S.B. Kang, Z.B. Xu, X, Tang, H.Y. Shum, Flash cut: foreground extraction with flash and no-flash image pairs, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, Minnesota, 2007.
[29] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, M. Cohen, Interactive digital photomontage, ACM Transactions on Graphics 23 (2004) 292–300.
[30] J. Sun, N.-N. Zheng, H.-Y. Shum, Stereo matching using belief propagation, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (7) (2003) 1–14.
[31] V. Kwatra, A. Schodl, I. Essa, G. Turk, A. Bobick, Graphcut textures: image and video synthesis using graph cuts, ACM Transactions on Graphics 22 (3) (2003) 277–286.
[32] Y.L. You, W.Y. Xu, A. Tannenbaum, M. Kaveh, Behavioral analysis of anisotropic diffusion in image processing, IEEE Transactions on Image Processing 5 (11) (1996) 1539–1553.
[33] G. Aubert, P. Kornprobst, Mathematical Problems in Image Processing—Partial Differential Equations and the Calculus of Variations, Springer-Verlag, New York, 2002.
[34] D. Donoho, Denoising by soft-thresholding, IEEE Transactions on Information Theory 41 (3) (1995) 613–627.

[35] C.L. Zitnick, N. Jojic, S.B. Kang, Consistent segmentation for optical flow estimation, in: Proceedings of the Tenth IEEE International Conference on Computer Vision, vol. 2, 2005, pp. 1308–1315.
[36] C. Liu, W.T. Freeman, R. Szeliski, S.B. Kang, Noise estimation from a single image, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), New York, USA, June 17–22, 2006, pp. 901–908.
[37] S.Z. Li, Markov Random Field Modeling in Image Analysis, Springer-Verlag, New York, 2001.
[38] J.W. Modestino, J. Zhang, A Markov random field model-based approach to image interpretation, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (6) (1992) 606–615.
[39] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (9) (2004) 1124–1137.
[40] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1222–1239.
[41] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient belief propagation for early vision, International Journal of Computer Vision 70 (1) (2006) 41–54.
[42] K. Murphy, Y. Weiss, M. Jordan, Loopy belief propagation for approximate inference: an empirical study, in: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden, 1999, pp. 467–475.
[43] M.J. Wainwright, T.S. Jaakkola, A.S. Willsky, MAP estimation via agreement on (hyper)trees: message-passing and linear-programming approaches, IEEE Transactions on Information Theory 51 (11) (2005) 3697–3717.
[44] V. Kolmogorov, Convergent tree-reweighted message passing for energy minimization, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (10) (2006) 1568–1583.
[45] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A comparative study of energy minimization methods for Markov random fields, in: Proceedings of European Conference on Computer Vision (ECCV), Graz, Austria, 2006, pp. 16–29.
[46] A. DeStefano, P.R. White, W.B. Collis, Training methods for image noise level estimation on wavelet components, EURASIP Journal on Applied Signal Processing 16 (2004) 2400–2407.
[47] B. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–608.
[48] E. Simoncelli, Statistical models for images: compression restoration and synthesis, in: Proceedings of Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, November 2–5, 1997, pp. 673–678.
[49] J. Portilla, V. Strela, M. Wainwright, E.P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, IEEE Transactions on Image Processing 12 (11) (2003) 1338–1351.
[50] S. Roth, M.J. Black, Fields of Experts: a framework for learning image priors, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, USA, June 20–25, 2005, pp. 860–867.
[51] M.F. Tappen, C. Liu, E.H. Adelson, W.T. Freeman, Learning Gaussian conditional random fields for low-level vision, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, USA, June 17–22, 2007, pp. 1–8.
[52] A. Buades, B. Coll, J.M. Morel, A non-local algorithm for image denoising, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, June 20–25, 2005, pp. 60–65.

About the Author—JIAN SUN received his bachelor's degree from University of Electronic Science and Technology of China in 2003, and Ph.D. in applied mathematics from Xi'an Jiaotong University in June 2009. He worked as a visiting intern at Microsoft Research Asia from November 2005 to March 2008, and postdoctoral research associate at computer vision lab. in University of Central Florida from August 2009 to March 2010. He now serves as a faculty at the School of Science, Xi'an Jiaotong University, China. His current research interests are image denoising, inpainting, segmentation and super-resolution.

About the Author—ZONGBEN XU received the M.S. degree in mathematics in 1981 and the Ph.D. degree in applied mathematics in 1987 from Xi'an Jiaotong University, China. In 1998, he was a postdoctoral researcher in the Department of Mathematics, The University of Strathclyde, United Kingdom. He worked as a research fellow in the Information Engineering Department from February 1992 to March 1994, the Center for Environmental Studies from April 1995 to August 1995, and the Mechanical Engineering and Automation Department from September 1996 to October 1996, at The Chinese University of Hong Kong. From January 1995 to April 1995, he was a research fellow in the Department of Computing in The Hong Kong Polytechnic University. He has been with the School of Science and Institute for Information and System Sciences at Xi'an Jiaotong University since 1982, where he was promoted to associate professor in 1987 and full professor in 1991, and now serves as an authorized Ph.D. supervisor in mathematics and computer science, director of the Institute for Information and System Sciences, and Vice President of Xi'an Jiaotong University. Currently, he serves as Chief Scientist of one National Basic Research Project of China (973 Project).