Pattern Recognition 46 (2013) 976–988

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Edge preserving image denoising with a closed form solution

Shifeng Chen [a], Ming Liu [b], Wei Zhang [a,*], Jianzhuang Liu [a,b,c]

[a] Shenzhen Key Lab for CVPR, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, PR China
[b] Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, PR China
[c] Media Lab, Huawei Technologies Co. Ltd., China

ARTICLE INFO

Article history:
Received 12 May 2011
Received in revised form
17 August 2012
Accepted 21 August 2012
Available online 2 September 2012

Keywords:
Image denoising
Markov random field
Label relaxation
Closed form solution

ABSTRACT

This paper addresses the problem of image denoising which is still a valid challenge at the crossing of functional analysis and statistics. We herein propose a novel pixel-based algorithm, which formulates the image denoising problem as the maximum a posterior (MAP) estimation problem using Markov random fields (MRFs). Such an MAP estimation problem is equivalent to a maximum likelihood (ML) estimation constrained on spatial homogeneity and is NP-hard in discrete domain. To make it tractable, we convert it to a continuous label assignment problem based on a Gaussian MRF model and then obtain a closed form globally optimal solution. Since the Gaussian MRFs tend to over-smooth images and blur edges, our algorithm incorporates the pre-estimated image edge information into the energy function construction and therefore better preserves the image structures. In the algorithm, patch similarity based pairwise interaction is also involved to better preserve image details and make the algorithm more robust to noise. Based on the theoretical analysis on the deviation caused by the discretization from obtained continuous global optimum to discrete output, we demonstrate the guaranteed optimal property of our algorithm. Both quantitative and qualitative comparative experimental results are given to demonstrate the better performance of our algorithm over several existing state-of-the-art related algorithms.

1. Introduction

Due to the defect of the image acquisition and transmission system, images are often corrupted by noise or even damaged to lose some pixel values. The contamination on images not only affects their visual quality but also precludes many further high level computer vision tasks such as image/video coding, recognition, scene understanding, and object tracking. Therefore, as an important image/video processing task, either as a stand-alone processing or as a pre-processing, it remains one of the most active topics in image processing, and draws much attention from the scholars all over the world. In most cases, noised images can be modeled as

$$X = F + N, \tag{1}$$

where X, F, and N represent the observed noisy image, the "true" image and the noise which is often the stationary Gaussian noise with zero mean and variance σ^2. The goal of image denoising is to remove the noise while maintaining and recovering the useful features and details of image as much as possible.

1.1. Related work

Over decades, a number of researchers dedicated their effort to the search for efficient image denoising approaches and many approaches were developed. They can be grouped into two basic classes: spatial filtering methods, which in essence estimate each pixel or region with neighboring pixels or regions in some way, and frequency domain filtering methods [1,2], whose basic idea is to project the image onto a set of orthogonal basis, usually referred to as wavelet basis, and then cut off the low coefficients (almost represent the noise) in the transformed representation using some kind of threshold or by shrinking. In this paper, we focus on the discussion and comparison of spatial filtering methods.

Traditional image denoising methods use linear local smoothing filters to do the work. The most common and simplest one is the *Gaussian filter* [3], which is linear and has the advantage of fast computation. However, since these linear filters are all based on the assumption of stationarity of the whole image that is always dissatisfied in real-world images, linear filters are not able to preserve edges and texture well, causing loss of image detail information. Nonlinear models, on the other hand, can handle edges better and reduce the blurring effect. Many nonlinear filters are based on partial differential equations (PDEs). The basic idea of PDEs based methods is to deform an image by a partial differential equation, and to obtain the expected denoising result by solving this equation.

* Corresponding author. Tel.: +86 755 86392199; fax: +86 755 86392073.
E-mail addresses: shifeng.chen@siat.ac.cn (S. Chen), mliu6@ie.cuhk.edu.hk (M. Liu), cndylan@gmail.com (W. Zhang), jz.liu@siat.ac.cn (J. Liu).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.08.014

The *total variation* (TV) *filter* [4], which leads to solving a second-order nonlinear PDE. Smooth regions are transformed into piecewise constant regions in the TV filter. This is good at preserving edges, but it also causes the staircase effect and loss of image structure and texture information. To avoid this effect, a fourth-order PDE filter is combined with the TV filter in [5] and some algorithms based on iterated total variation [6,7] have been developed. Another popular model of PDEs based approaches is the *anisotropic filter* [8,9] which uses an anisotropic diffusion (AD) equation to smooth a noisy image for denoising. The strategy of AD is to impose smoothness within the regions achieved through pre-estimation, whereas less smoothness across edges for edge sharpness preserving. On the other hand, AD can be regard as a gradient descent on some energy function incorporating with local statistic characteristics [8]. The *bilateral filter* (SUSAN filter) [10,11] is also a commonly used nonlinear filter, which is different from the above local smoothing filters by the different use of spatial neighborhoods. It takes the average of the values of pixels which are close to each other in both grey levels and spatial locations, while the other local filters only consider pixels' geometric closeness. The bilateral filter can keep image edges relatively sharp. Furthermore, considerable interests have been given to the use of segmentation for image denoising [12–15]. These approaches share the same basic remark: smoothing the reference pixel or region by using the ones belonging to the same cluster which is obtained through the segmentation procedure and therefore preserving the edges well.

Recently, Buades et al. [16,17] presented a non-local means (NL-means) algorithm for image denoising. They argued that the local smoothing methods aim at noise reduction and the reconstruction of the main geometrical configurations of the image, but not at the preservation of fine structures, details, and textures. To address this problem, the NL-means algorithm estimates the denoised value for a pixel as the weighted average of the values of all pixels whose Gaussian neighbors look like the neighbors of the pixel, that is, who has a close neighborhood configuration, based on the regularity assumption that self-neighborhoods similarities of each pixel exist in nature image. It is obvious that NL-means algorithm is suited for denoising the images with periodicity texture patterns. However, it is limited in many real-world images that fail in the assumption or in the worse noise distorted images due to the corruption of the structures.

One of the best state-of-the-art denoising methods in the spatial domain is called Block-Matching and 3D Filtering (BM3D) [18]. This method is based on an enhanced sparse representation in transform-domain. The enhancement of the sparsity is achieved by grouping similar 2D image blocks into 3D data arrays called "groups". More recent denosing algorithm proposed by Bouboulis et al. [19] is based on the theory of Reproducing Kernel Hilbert Space (RKHS). The main advantage of this algorithm is that it can remove any kind of noise from images, in contract to the mostly used noise dependent denoising methods. From the comparative results, BM3D obtains the better results for Gaussian noise.

Thanks to the success and rapid advances of the Markov random field (MRF) models, which provide a robust and unified framework for many early computer vision tasks, some image denoising methods also formulate the problem as an energy minimization problem using MRFs [20]. This formulation is justified in terms of maximum a posteriori (MAP) estimation of a Markov random field in the Bayesian framework, since the true image is regarded as the realization of a Markov random field. To the best of our knowledge, for the general case, the energy minimization problem in discrete domain is NP-hard. The major obstacle of the optimization is the large computational cost owing to the high dimensional computing space. Many approaches have been proposed to solve the MRFs energy optimization. Simulated annealing, popularized in [20], can minimize an arbitrary energy function but is very slow. Iterated conditional modes (ICM) introduced in [21], iterates greedily to a local minimum sensitive to initial value. Recent developed algorithms approximately solve MRFs by graph cuts [22–24] or belief propagation [25,26] with the consideration about the tradeoff between accuracy and efficiency. Graph cuts and belief propagation both make it in discrete sense, and thus usually can only find a locally optimal solution to the problem, which depends on the initialization and the convergence criteria. Moreover, they carry out the denoising task iteratively and do not have a closed form solution. We also found that the methods using graph cuts and belief propagation cannot well preserve image edges and the large labels in denoising task, especially for color image, cause denoising procedure to be slow.

1.2. Overview of the paper

In this paper, we also formulate image denoising with an MAP-MRF model. However, we solve the MAP problem by transforming it into a continuous optimization problem where pixels' intensities are relaxed from discrete values to continuous values. Compared with the related approaches, our proposed novel denoising algorithm has the following merits.

(1) Without any approximation in the deduction process, a closed-form globally continuous optimal solution can be obtained directly, which provides the good prerequisite for our final denoising result.
(2) Image edge details can be better preserved in our algorithms since the pre-estimated edge information is incorporated into our object MRF energy function and the special design for the smoothness weight imposes the reasonable smoothness constraint conditioned on the observed intensities.
(3) Our formulation for the gray level image denoising can directly extend to operating on the color image in the CIE-Lab color space uniformly at once and hence does not need more computational complexity in contrast with the much larger computational cost as the increase of the label number in belief propagation and graph cuts algorithms. Since CIE-Lab color approximates human vision and is perceptually uniform, our denoising algorithm can keep the balance of colors and avoid the unexpect color combinations after denoising operation, which may appear in many denoising algorithms due to filtering the color bands separately.
(4) Further to note that with the construction of the undirect weighted graph representing the energy function, we utilize the Laplacian matrix corresponding to the graph to speedup our algorithm and the final closed form solution is also formulated in this way.
(5) It is inevitable that an energy deviation from continuous global optimum is generated by discretization for each pixel's value which is essential for image output. Fortunately, based on the theoretical analysis, we obtain an upper bound of the deviation which guarantees the optimal property of our algorithm. Finally, experimental results demonstrate all the advantages of our approach and show that our method outperforms the related methods quantitatively and qualitatively.

This paper is the extended version of our work in [27]. Compared to [27], we review more related work, give Optimal Property of our algorithm, and report much more experimental results.

2. MAP-MRF model

Markov Random Fields were first introduced in vision in [20]. In the context of the image denoising problem, we represent the

components of MRFs with $X=[x_1,x_2,\ldots,x_n]^T$ and $F=[f_1,f_2,\ldots,f_n]^T$, where X and F denote the observed image with noise and the denoising image, respectively, and n is the number of pixels of the image. Here x_i and f_i regarded as random variables denote the intensity values of pixel i, $1\le i\le n$.

In this paper, we focus on grey-level images, but the formulation can be directly extended to handling color images. We will demonstrate this extension in the relevant point below.

We formulate image denoising as an optimization problem to find an F that maximizes the posterior probability $P(F|X)$ given X which results in the globally optimal solution from a statistical viewpoint. With the Markov assumption that each random variable f_i only depends on the random variables in F that correspond to the neighborhood pixels of pixel i, each concrete F denotes a realization of the Markov random field. By using the Bayes' rule, we have

$$\begin{aligned}\underset{F}{\operatorname{argmax}}P(F|X)&=\underset{F}{\operatorname{argmax}}P(X|F)P(F)/P(X)\\&=\underset{F}{\operatorname{argmax}}P(X|F)P(F),\end{aligned}\tag{2}$$

where $P(X)$ can be removed since it is a constant with respect to F. Therefore, the MAP estimation is to find $\widehat{F}$ such that

$$\widehat{F}=\underset{F}{\operatorname{argmax}}P(X|F)P(F).\tag{3}$$

Here $P(X|F)$ is the likelihood function which can be represented by the sensor noise model. Assume that the pixel intensities follow an independent identical distribution (i.i.d.) and the noise model is the additive Gaussian white noise. Then $P(X|F)$ can be defined as

$$P(X|F)\propto\prod_{i=1}^{n}\exp(-d(x_i,f_i)),\tag{4}$$

where $\exp(-d(x_i,f_i))$ is the sensor noise model and $d(x_i,f_i)$ is called the data penalty function that penalizes the inconsistency between output and observed data, i.e., represents the penalty of x_i having the output f_i after denoising. In this paper, it is defined as

$$d(x_i,f_i)=|f_i-x_i|^2.\tag{5}$$

Concerning the MRFs, the Hammersley–Clifford theorem [21] states that the prior probability of a Markov random field is proportional to the summation over the cliques in all neighborhood. We model the prior probability $P(F)$ by MRFs whose clique potentials involve pairs of neighboring pixels, which is defined as

$$P(F)\propto\exp\left(-\sum_{i=1}^{n}\sum_{j\in\mathcal{N}(i)}s(f_i,f_j)\right),\tag{6}$$

where $\mathcal{N}(i)$ is the neighborhood of i, and $s(f_i,f_j)$ is the clique potential, which is also commonly called smoothness penalty function to imposing the spatial homogeneity. There have been some clique potential formats to enforce the intensity smoothness constraint for neighboring pixels, e.g., generalized Potts Model, $w_{ij}\cdot(1-\delta(f_i-f_j))$, ($\delta(\cdot)$ is the unit impulse function), and linear clique potential, $w_{ij}\cdot|f_i-f_j|$ [28]. The generalized Potts Model penalizes any neighborhood pair of different labels while the penalization is independent upon the difference. By contrary, the cost of spatial incoherency in clique potential is proportional to the difference between neighboring pixels' labels. In our approach, the clique potential is computed by

$$s(f_i,f_j)=w_{ij}\cdot|f_i-f_j|^2.\tag{7}$$

We herein choose quadratic label difference since it will facilitate to achieve the closed form solution and to speedup our algorithm running based on the usage of the Laplacian matrix. The importance of this formulation will be seen in the deduction of the closed form solution below. Meanwhile, the factor w_{ij} denotes the affinity value between pixels i and j and is used both to control the smoothness degree for each neighboring pair and to preserve edge sharpness based on the local statistical information and the pre-obtained edge localization information.

We design w_{ij} in this way: (1) if the intensity difference $\Delta(i,j)=|x_i-x_j|^2$ is large between the neighboring pixels i and j in the input image, the smoothness penalty $s(f_i,f_j)$ should be small; (2) the further distance between pixels i and j, the less effect of them on $s(f_i,f_j)$; (3) if pixels i and j fall into two regions separated by an edge, they have no effect on $s(f_i,f_j)$, i.e., $w_{ij}=0$. Based on these targets, we define w_{ij} as

$$w_{ij}=a\cdot\exp\left(-\frac{\Delta(i,j)}{b}\right)\cdot k(i,j)\cdot T(C_i=C_j),\tag{8}$$

where a and b are two positive factors to control the contribution of w_{ij} to $s(f_i,f_j)$, $k(i,j)=\exp(-\|i-j\|^2/2)$ is a Gaussian kernel function to reach the above target (2), and $T(C_i=C_j)$ is used to reach the above target (3). The term $\exp(-\Delta(i,j)/b)$ has been first used to control smoothness in [29] for the stereo vision problem. $T(\cdot)$ is 1 if its argument is true and 0 otherwise. C_i and C_j are region indexes that can be explained with Fig. 1. In the figure, an edge separates the window into two regions and pixels i and j_1 have the same region index but pixels i and j_2 have different region indexes, i.e., $C_i=C_{j_1}$ and $C_i\neq C_{j_2}$. We use Canny edge detector to find edges in the input image since this edge detector is insensitive to noise, and then assign indexes to different regions. If there is no edge detected via the Canny detector in a noisy image, then the denoising result can also be obtained with $T(C_i=C_j)=1$ for all pixel pairs (i,j) in (8). We further exhibits the intuitional representations of w_{ij} images of the left, up, left up, and left down neighbor weights for all the pixels of the "House" image, respectively, in Fig. 2. The visual w_{ij} output clearly indicates our motivation of the design in (8), that is, smoothing constraints are only imposed on the homogeneous areas while allowing discontinuities and well demonstrates our algorithm's merit that the design of w_{ij} can ensure the image edges be well preserved.

From (3), (4) and (6), we can see that the MAP estimation is equivalent to minimizing the following energy function:

$$E(F)=\sum_{i=1}^{n}d(x_i,f_i)+\sum_{i=1}^{n}\sum_{j\in\mathcal{N}(i)}s(f_i,f_j).\tag{9}$$

The first term on the right-hand side is called the data term and the second one is the smoothness term. It is obvious that the MAP estimate F which minimizes the energy $E(F)$, that is, the minimal cost realization of a Markov random field, tends to be consistent

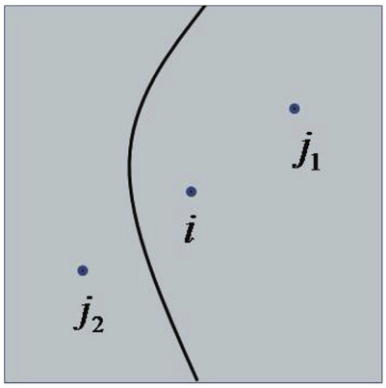

Fig. 1. Explanation of the region indexes. The curve denotes an edge separating the window into two regions. Pixels i and j_1 have the same region index ($C_i=C_{j_1}$), but pixels i and j_2 have different region indexes ($C_i\neq C_{j_2}$).

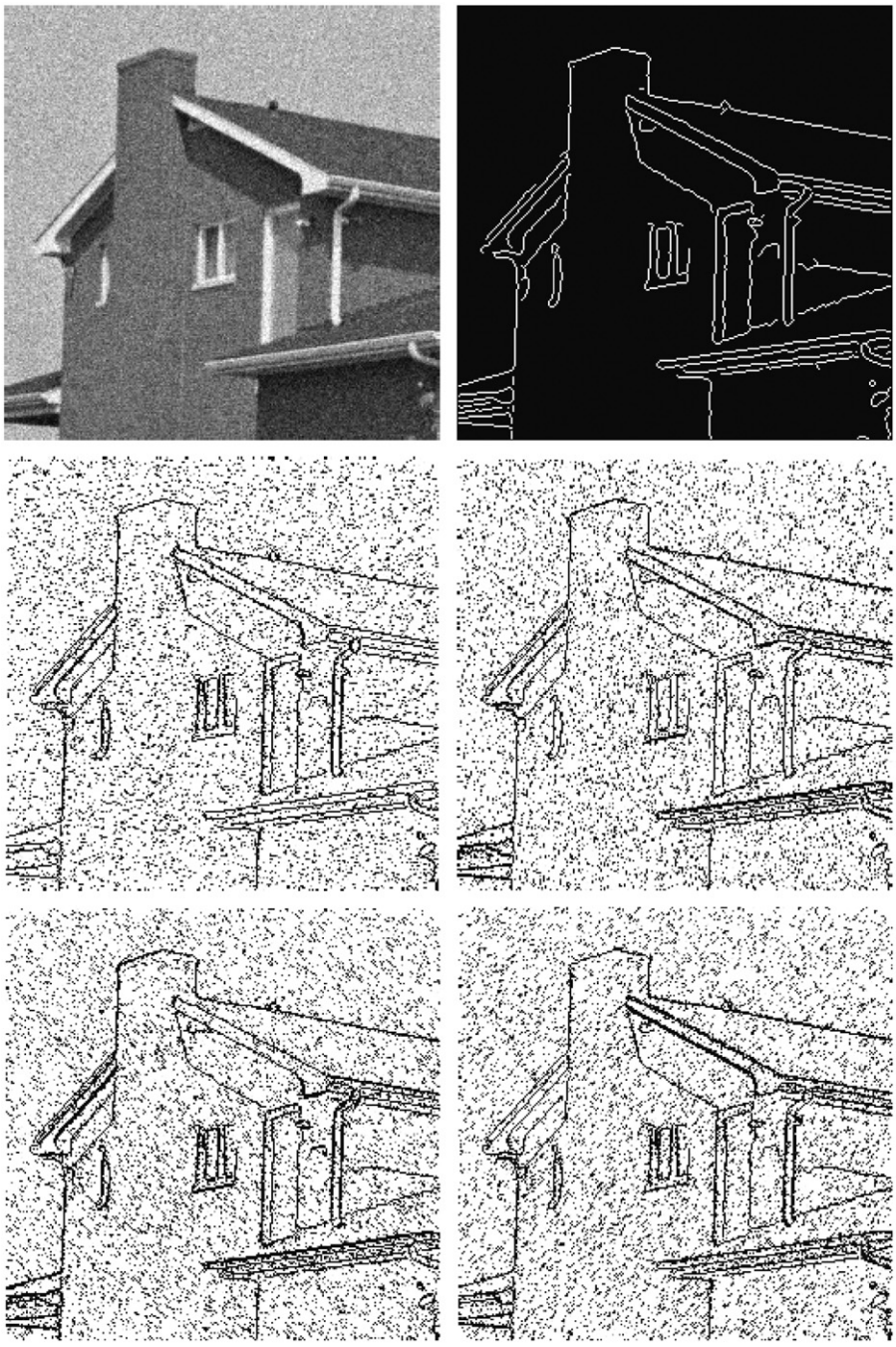

Fig. 2. Visual representation for w_{ij} of noised "House" in 8-connected neighborhood system. From left to right and from top to down: noised "House" image with Gaussian noise $\sigma = 20$, corresponding edge image and w_{ij} of each pixel with its left, up, left up and left down neighbor, respectively.

with the observed data X as well as containing small number of discontinuities as much as possible.

The approaches in [22–26] solve the energy minimization problem by graph cuts and belief propagation, respectively, with their own energy functions in a similar form to (9). However, we pay more attention to preserve edges when designing our energy function. On the other hand, since such an optimization problem is NP-hard, the solutions obtained by graph cuts and belief propagation are locally optimal. In the next section, we relax the pixel intensities from discrete values to continuous values and meanwhile introduce the utilization of Laplacian matrix corresponding to our constructed undirected weighted graph that makes us easily get the globally optimal closed form solution and solve it faster. Another advantage of such a continuous model is that we can give the theoretical analysis for the optimal properties (see Section 4).

In the anisotropic diffusion (AD) model [8], the image is updated iteratively by

$$\frac{\partial I_t(x,y)}{\partial t} = \mathbf{div}[c_t \cdot \nabla I_t(x,y)], \tag{10}$$

where $I_t(x,y)$ is the gray level at coordinates (x,y) of the image at iteration t, **div** represents the divergence operator, $\nabla I_t(x,y)$ is the gradient of the image, and c_t represents the diffusion coefficient. For iteration 0, the image is the input noisy image, i.e., $I_0 = X$; after T iterations, the denoising result $F = I_T$. To preserve image detail information, the diffusion coefficient c_t can be defined in the forms combining image edge information, for example in [30].

In our algorithm, image edge information is incorporated in the regularization term (the smoothness term). Different from the AD model, our MAP-MRF model obtains the final results with a closed-form solution. This avoids the selection of iteration number T, which is important for the result of the AD approaches.

3. A closed form solution

From the definitions of the data penalty and smoothness penalty functions in (5) and (7), the energy function can be written as

$$E(F)=\sum_{i=1}^{n}|f_i-x_i|^2+\sum_{i=1}^{n}\sum_{j\in\mathcal{N}(i)}w_{ij}\cdot|f_i-f_j|^2. \tag{11}$$

With the smooth term in (11), our proposed MRF is isotropic obviously and we construct an undirect weighted graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ where $\mathcal{V}$ is the set of vertices denoting the image pixels, and $\mathcal{E}$ is the set of weighted edges. Then the adjacent matrix of $\mathcal{G}$ is $W=[W_{ij}]_{n\times n}$ whose elements are defined as

$$W_{ij}=\begin{cases} w_{ij} & \text{if } i\neq j, j\in\mathcal{N}(i),\\ 0 & \text{if } i\neq j, j\notin\mathcal{N}(i),\\ c & \text{if } i=j,\end{cases} \tag{12}$$

where $c>0$ is some constant, which indicates the constant weight edge between vertex i and itself. It is explicit that these extra edges correspond to the terms $c\cdot(f_i-f_i)$, $1\le i\le n$ which have no effect on the energy function and its optimization. Let D be an $n\times n$ diagonal matrix with the (i,i)-th entry $D_{ii}=\sum_{j=1}^{n}W_{ij}$. If $D_{ii}=0$, then we say that i is an isolated vertex in the graph which leads W to be singular. The constant c in the (i,i)-th entry of W can avoid this singularity. Moreover, the importance of constant c builds up the numerical stability of our solution. With the design of w_{ij} in (8), D_{ii} is very small for some pixels that have a few edges connecting to their neighborhood or only have small weighted edges. Since our deduction for the closed form solution below involves D^{-1} and $D^{-1/2}$, the proper c can make those matrices numerically stable, that is, important for us to achieve the accurate globally optimal solution. We also find that it is better for the constant c to be comparable with the value of parameter a in (8). In our algorithm, we choose $c=a$.

With the construction of energy function $E(F)$ and corresponding graph $\mathcal{G}$, the following theorem gives the insight of our closed form solution to the utilizing the Laplacian matrix with continuous labels relaxation.

Theorem 1. *Let $X=[x_1,x_2,\ldots,x_n]^T$ and $F=[f_1,f_2,\ldots,f_n]^T$. W_{ij} is defined in (12), where f_i, $1\le i\le n$, are continuous variables. Then $D^{-1/2}(D^{-1}+2\bar{L})^{-1}D^{-1/2}X$ is the closed form solution which is globally optimal for the below energy function minimization*

$$E(F)=\sum_{i=1}^{n}|f_i-x_i|^2+\sum_{i=1}^{n}\sum_{j\in\mathcal{N}(i)}W_{ij}\cdot|f_i-f_j|^2, \tag{13}$$

where $\bar{L}$ is the normalized Laplacian matrix of $\mathcal{G}$ constructed above and is expressed as

$$\bar{L}=D^{-1/2}(D-W)D^{-1/2}. \tag{14}$$

Proof. Let $R=[r_1,r_2,\ldots,r_n]^T$, where $r_i=\sqrt{D_{ii}}\cdot f_i$, $1\le i\le n$, form a set of medium variables. Then we have

$$F=D^{-1/2}R, \tag{15}$$

and the energy function in (13) can be written as

$$E(F)=\overline{E}(R)=\sum_{i=1}^{n}\left|\frac{r_i}{\sqrt{D_{ii}}}-x_i\right|^2+\sum_{i,j=1}^{n}W_{ij}\cdot\left|\frac{r_i}{\sqrt{D_{ii}}}-\frac{r_j}{\sqrt{D_{jj}}}\right|^2, \tag{16}$$

which has a compact matrix form expression:

$$\overline{E}(R)=\|D^{-1/2}R-X\|^2+2(R^T\bar{L}R), \tag{17}$$

where

$$\|D^{-1/2}R-X\|^2=\sum_{i=1}^{n}\left|\frac{r_i}{\sqrt{D_{ii}}}-x_i\right|^2, \tag{18}$$

$$2(R^T\bar{L}R)=\sum_{i,j=1}^{n}W_{ij}\cdot\left|\frac{r_i}{\sqrt{D_{ii}}}-\frac{r_j}{\sqrt{D_{jj}}}\right|^2. \tag{19}$$

To minimize $E(R)$, taking its derivative with respect to R and setting it to zero yields

$$\frac{\partial\overline{E}(R)}{\partial R}=2D^{-1/2}(D^{-1/2}R-X)+4\bar{L}R=0, \tag{20}$$

which results in

$$(D^{-1}+2\bar{L})R=D^{-1/2}X. \tag{21}$$

Since W and $\bar{L}$ are positive semi-definite, $(D^{-1}+2\bar{L})$ is nonsingular. Therefore, a closed form solution for the medium variable R is obtained as

$$R=(D^{-1}+2\bar{L})^{-1}D^{-1/2}X. \tag{22}$$

Thus the globally optimal solution is

$$F=D^{-1/2}R=D^{-1/2}(D^{-1}+2\bar{L})^{-1}D^{-1/2}X. \qquad \square \tag{23}$$

Similarly, for color images, we treat each color as a point in a three dimensional space, i.e., vector with three components, underlying the CIE-Lab color space. Based on the perceptually linearity of CIE-Lab, we can use Euclidean distance in L*a*b space as color difference measurement used in (8) which is consistent with human perception. With the replacement of gray level intensity value as color vector, we have the uniform closed form solution for color image denoising as (23), since L*a*b three channels are relatively independent and share the same W_{ij}. Hence, our algorithm does not need more computational cost for color image denoising despite the larger labels.

It is worth noting that above derivation of globally optimal F is based on the assumption that R and F are continuous variables (see (15) and (20)). In the context of image denoising, however, pixel intensities are discrete. Therefore, after we obtain the continuous solution, we should make discretization which naturally results in the deviation of denoising output from the globally optimal solution. Fortunately, we find that discretization result slightly falls away and the error is even within a known bound. The details of the optimal property will be elaborated in the next section.

4. Optimal property

Optimal property discussion is meaningful for MRFs energy optimization techniques, especially for those that converge to a local minimum in some criteria, e.g., iterated conditional modes (ICM) [21], graph cuts based expansion algorithm and swap algorithm [22] and belief propagation [25,26]. To the best of our knowledge, except for the expansion algorithm that is within a known factor $2c$ of the global optimum, where $c=\max_{i,j\in\mathcal{N}}(\max_{f_i\neq f_j}s(f_i,f_j)/\min_{f_i\neq f_j}s(f_i,f_j))$, other algorithms converging to local minimum do not have theoretical analysis for the optimal

properties, although it is undeniable that many of them have desirably visual outputs.

As discussed above, we relax the discrete pixels' intensities to continuous ones and therefore achieve the globally optimal solution in continuous domain in our algorithm. To present the output image, discretization is naturally needed. We herein adopt the rounding operation to make discretization. It is straightforward that the error caused by discretization of intensity values affects the output energy described in (11). Although in some cases the discretization may make final energy even lower, which is better for our denoising, we should consider the worst case to give an insight into the optimal property of our approach. Fortunately, the error between globally optimal energy and the one after discretization is within a known bound described in the following theorem, and therefore our algorithm has the guaranteed optimality properties.

First, we introduce some notations. Let f_i^* and f_i denote the optimal continuous label and corresponding discretized one for pixel i, $1 \le i \le n$, where n is the number of pixels of input image. E_{opt}, E_{dis} represent the globally optimal energy in continuous domain and the energy after label discretization, respectively. Furthermore, we have $E_{opt} = E_{data} + E_{smooth}$, where $E_{data} = \sum_{i=1}^{n} (f_i^* - x_i)^2$ and $E_{smooth} = \sum_{i=1}^{n} \sum_{j \in \mathcal{N}(i)} w_{ij} \cdot (f_i^* - f_j^*)^2$. With these notations, the following theorem demonstrates the energy error bound between E_{opt}, E_{dis}.

Theorem 2. *The error between E_{opt} and E_{dis}, represented by $\Delta E = E_{dis} - E_{opt}$, holds the upper bound $2^{(1/2)(\lceil \log_2^{4(1+k)n} \rceil)} \sqrt{cE_{opt}} + A$, that is,*

$$\Delta E \le 2^{(1/2)(\lceil \log_2^{4(1+k)n} \rceil)} \sqrt{cE_{opt}} + A, \tag{24}$$

where $A \sim \mathcal{O}(n)$, is some constant, k is the number of neighborhood for each pixel and $c = \max(\max_{i,j}(w_{ij}), 0.5)$. Meanwhile, the operation $\lceil * \rceil$ denotes taking upper integer.

Proof. With the definition of f_i^*, f_i and energy function (11), we have

$$\begin{aligned}
\Delta E &= \sum_{i=1}^{n} \Big[(f_i - x_i)^2 - (f_i^* - x_i)^2 \\
&\quad + \sum_{j \in \mathcal{N}(i)} w_{ij} \cdot ((f_i - f_j)^2 - (f_i^* - f_j^*)^2) \Big] \\
&\le 2x \sum_{i=1}^{n} \left(|f_i^* - x_i| \cdot |d_i| + \sum_{j \in \mathcal{N}(i)} w_{ij} \cdot |f_i^* - f_j^*| \cdot |d_i + d_j| \right) \\
&\quad + \sum_{i=1}^{n} d_i^2 + \sum_{j \in \mathcal{N}(i)} w_{ij} \cdot (d_i + d_j)^2,
\end{aligned} \tag{25}$$

where $d_i = f_i - f_i^*$ for $1 \le i \le n$. It is obvious that d_i has the range $(-0.5, 0.5]$ due to the rounding operation for discretization and therefore $d_i + d_j \in (-1, 1]$.

The pairwise constraints result in that the maximum of $|d_i|$, and $|d_i + d_j|$, that is, 0.5 and 1, cannot be achieved simultaneously in all i, $1 \le i \le n$ and (i,j), $(i,j) \in \mathcal{N}$ unless the 4-connected neighborhood system in special cases. Without the consideration of the pairwise constraints we take $|d_i| = 0.5$, and $|d_i + d_j| = 1$ for all terms in (25) that is the worst situation for the deviation of E_{dis}. Then we have

$$\Delta E \le \sum_{i=1}^{n} \left(|f_i^* - x_i| + \sum_{j \in \mathcal{N}(i)} 2w_{ij} \cdot |f_i^* - f_j^*| \right) + A, \tag{26}$$

where $A = n/4 + \sum_{i=1}^{n} \sum_{j \in \mathcal{N}(i)} w_{ij}$, is a constant and $A \sim \mathcal{O}(n)$.

Based on the inequation $\sqrt{a} + \sqrt{b} \le \sqrt{2(a+b)}$ where $a, b \ge 0$, (26) can be further advanced as

$$\begin{aligned}
\Delta E &\le 2c \sum_{i=1}^{n} \left(\frac{1}{\sqrt{2c}} |f_i^* - x_i| + \sum_{j \in \mathcal{N}(i)} \sqrt{\frac{w_{ij}}{c}} |f_i^* - f_j^*| \right) + A \\
&\le 2c(\sqrt{2})^{(\lceil \log_2^{4(1+k)n} \rceil)} \sqrt{\frac{1}{2c} E_{data} + \frac{1}{c} E_{smooth}} + A \\
&\le 2^{(1/2)(\lceil \log_2^{4(1+k)n} \rceil)} \sqrt{cE_{opt}} + A,
\end{aligned} \tag{27}$$

where $c = \max(\max_{i,j}(w_{ij}), 0.5)$. □

Conclusively, we obtain the upper bound for ΔE which demonstrates the guaranteed optimal property of our algorithm, that is to say, our algorithm achieves an energy close to the global optimum.

5. Experimentation

To demonstrate the performance of our algorithm, we compare our algorithm with five related algorithms: graph cuts (GC) [22–24], belief propagation (BP) [25,26] representing the MRFs based approaches, Gaussian filter (GF) [3] which is the representation of the linear local smoothing filters, bilateral filter (BF) [11] which represents the nonlinear local smoothing filters, and NL-means (NL) [16] which is a nonlocal approach. We test these algorithms on a set of classic gray scale images, "Barbara", "Boat", "House", "Pepper", "Lena" of size 256×256 pixels, and the Berkeley segmentation benchmark [31] containing 300 natural images, which are all contaminated with five different noise level additive Gaussian white noise (standard deviation $\sigma = 10$, 20, 30, 50 and 100). At last, we test our algorithm on image "Lena", "Barbara", and "Boat" with the size of 512×512 to compare with the BM3D [18] method. Both visual quality comparisons and quantitive comparisons, which use peak signal-to-noise ratio (PSNR) and mean squared error (MSE), are showed. Here all the algorithms are implemented in Matlab running on a Pentium IV PC (2.8 GHz CPU).

We arrange the denoising output comparisons with two sections, comparison with MAP-MRFs model based algorithms and with related state-of-art denoising algorithms. The experiments indicate the better performance of our algorithm both among the algorithms using same model and other well known approaches.

5.1. Comparison with model variants

Belief propagation and graph cuts are two main optimization techniques for MAP-MRFs model, and thus have been developed constantly in recent years. As a same model based approach, we compare our denoising results with the two mature approaches to better demonstrate superiority and efficiency of our algorithm.

Under the MAP-MRF models, one of the most important thing, which greatly affects the final results, is the object energy function design. As indicated above, our algorithm focuses on the quadratic data and smooth term with special design w_{ij} for different neighborhoods to allow intensity discontinuity on the image edges and smoothness in homogenous regions. We herein choose best energy function formats for belief propagation and graph cuts in terms of best visual denoising quality to make our comparison more convincing. The energy function for GC and BP denoising consists of truncated quadratic data term and truncated quadratic smooth term with spatially constant parameter λ to control the tradeoff between data and smooth constraint. The truncation on data penalty and smooth penalty can handle the outlier noise and preserve the intensity discontinuity, respectively.

Some results on the noisy images "Barbara" and "Boat" with standard deviation $\sigma = 20$, and 30 are provided in Figs. 3 and 4 for visual comparison. The figures indicate that the results of our algorithm look better than the others and the edges are well preserved although some small details are lost. The details and main structures are removed in the seriously over-smooth results of BP and GC, especially in the seriously corrupted images. Due to the tradeoff consideration between data truncation constant and smoothing degree, there are also some isolated undesirable noise pixels in GC and BP results which badly affects the image visual quality. However, outlier noise pixels are all removed in our approach results although we do not use data truncation.

The PSNR and MSE values on "Barbara", "Boat" denoising with five different level noise are given in Table 1. From the

Fig. 3. Results of MRFs-based denoising algorithms on "Barbara" image with the noise $\sigma = 20$ (the first row) and $\sigma = 30$ (the second row). From left to right: the noisy image, the result of BP, GC, and our algorithm.

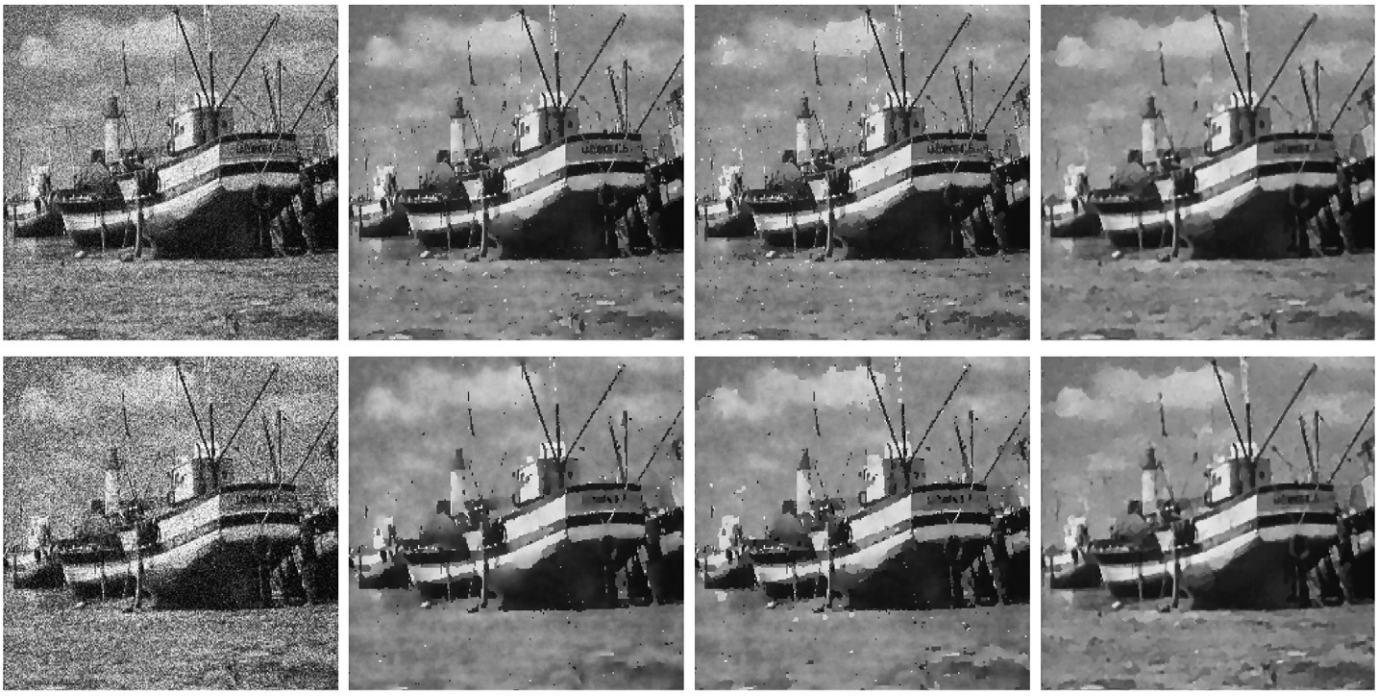

Fig. 4. Results of MRFs-based denoising algorithms on "Boat" image with the noise $\sigma = 20$ (the first row) and $\sigma = 30$ (the second row). From left to right: the noisy image, the result of BP, GC, and our algorithm.

quantitative comparison, we can see that our algorithm outperforms BP and GC. The output PSNR values of our algorithm are all higher than that of the other two. The improvement is little in low level noisy images and great in worse corrupted ones and the BP algorithm obtains better PSNR results than GC. On the other hand, our approach is also on top of the other two in terms of MSE.

Table 1
PSNR and MSE comparison among MRFs-based denoising algorithms: Graph Cuts(GC), Belief Propagation(BP) and Ours, on the noisy images "Barbara" and "Boat" with five different noise level. $PSNR_0$ is the PSNR value of the noisy image.

Image	σ/$PSNR_0$	PSNR			MSE		
		GC	BP	Ours	GC	BP	Ours
Barbara	10/28.13	33.77	33.88	**34.14**	27.31	26.62	**25.09**
	20/22.18	31.29	31.50	**31.54**	48.34	46.03	**45.64**
	30/18.61	30.51	30.81	**31.16**	57.78	53.99	**49.77**
	50/14.15	29.67	30.17	**30.28**	70.13	62.6	**61.00**
	100/8.14	27.24	27.77	**28.30**	122.89	108.62	**96.20**
Boat	10 / 28.14	34.29	**34.36**	34.20	24.22	**23.85**	24.72
	20/22.08	32.02	32.22	**32.32**	40.84	38.97	**38.09**
	30/18.64	31.21	31.60	**31.89**	49.26	44.95	**42.06**
	50/14.13	29.79	30.68	**30.81**	68.25	55.62	**54.01**
	100/8.13	27.44	28.00	**28.44**	117.37	103.05	**93.04**

Table 2
Optimal property of our algorithm on the five noisy images with standard deviation $\sigma = 20$.

Image		Barbara	Boat	House	Pepper	Lena
Global optimum	$E_{opt}(10^7)$	2.4309	2.4086	0.9128	1.8740	1.3662
Output energy	E (10^7)	2.4353	2.4133	0.9205	1.8798	1.3730
Theoretical	*bound* (10^6)	7.8215	7.7855	6.1875	6.8674	6.9791
	factor $\sim$	1.4	1.4	1.7	1.4	1.6
Practical	*error* (10^6)	0.044	0.047	0.077	0.058	0.068
	factor $\sim$	1.0018	1.0020	1.0084	1.0031	1.0050

It should be mentioned how we choose the parameters for the algorithms. We found the best parameters in terms of the largest PSNR for each algorithm with respect to each noisy image, which means that each algorithm runs with five sets of parameters to denoise the five noisy images, respectively. In our algorithm, the parameter $b=100$ is fixed and the other parameter a ranges from 1 to 3. In GC and BP algorithms, data truncation constant, smooth truncation constant and λ range from 3000 to 10000, 200 to 500 and 2 to 10, respectively, while the iteration number is 15 to ensure the convergence of BP and GC. We also find that the parameter setting with respect to the noise level is stable and robust in our algorithms for all the images whereas BP and GC are not due to the strong dependence on the initial estimation and large ranges of parameters. Obviously, the less parameter and the parameter stability in our algorithm make us convenient to achieve the best denoising results.

It is also worth noticing that our algorithm can obtain the closed form solution faster than iterative minimization based GC and BP, even when the efficient BP proposed in [26] is used. For 256×256 noisy image, the computational time taken by the GC, BP, and our algorithms are about 3 min, 1 min, and 5 s, respectively.

As energy minimization techniques, graph cuts and belief propagation converge to a strong local minimum. Meanwhile, the deviation from globally continuous optimum to discrete output in our algorithm is unavoidable. Therefore, optimal property in terms of deviation from global optimum is important. We know that the GC expansion algorithm produces a solution within a known factor $2c$ of the global optimum, where $c = \max_{i,j \in \mathcal{N}} (\max_{f_i \neq f_j} s(f_i,f_j)/\min_{f_i \neq f_j} s(f_i,f_j))$. The factor can be as small as 2 in the best situation where smooth term is Potts model, but Potts model does not suit for the denoising task since Potts model denoising generates serious staircase effect. Under other smooth constraint models applicable for expansion algorithm like linear truncated one, the optimal factor is always large in image denoising, which leads to the insignificant optimal guarantee. Fortunately, with our energy function and solving technique,

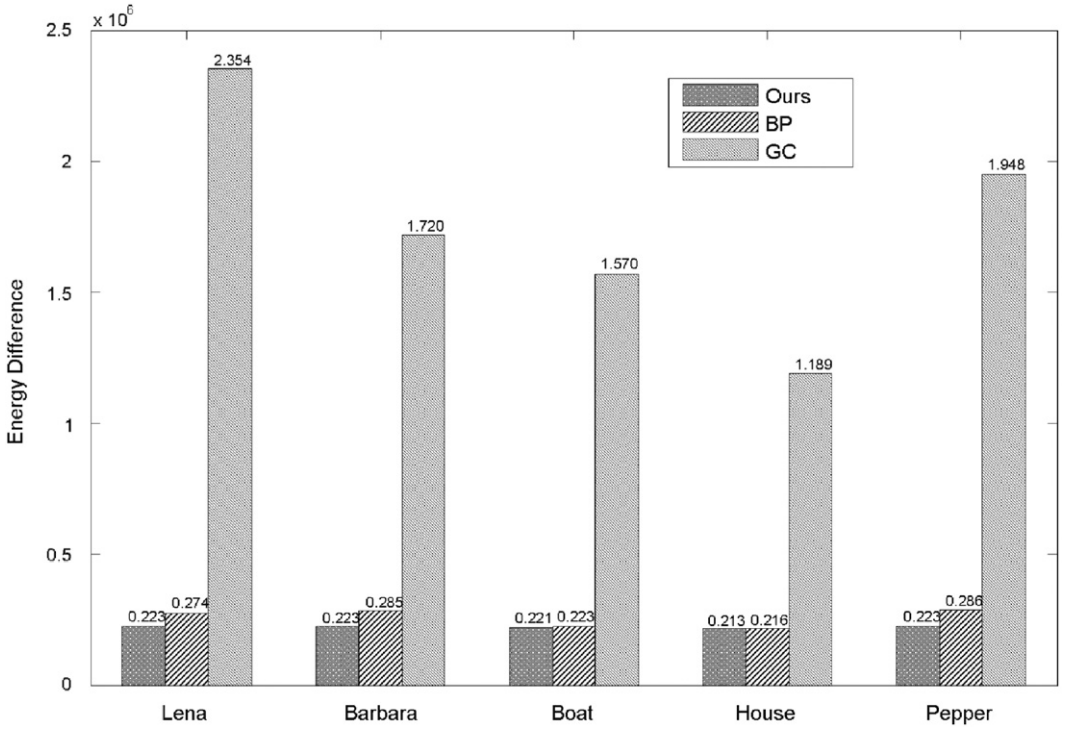

Fig. 5. Energy difference between continuous global optimum and the energy of the GC, BP and Our denoising outputs using the same MRFs energy format and parameters on the five test images with the noise $\sigma = 20$, respectively. In each group, from left to right: GC energy, BP energy and Ours.

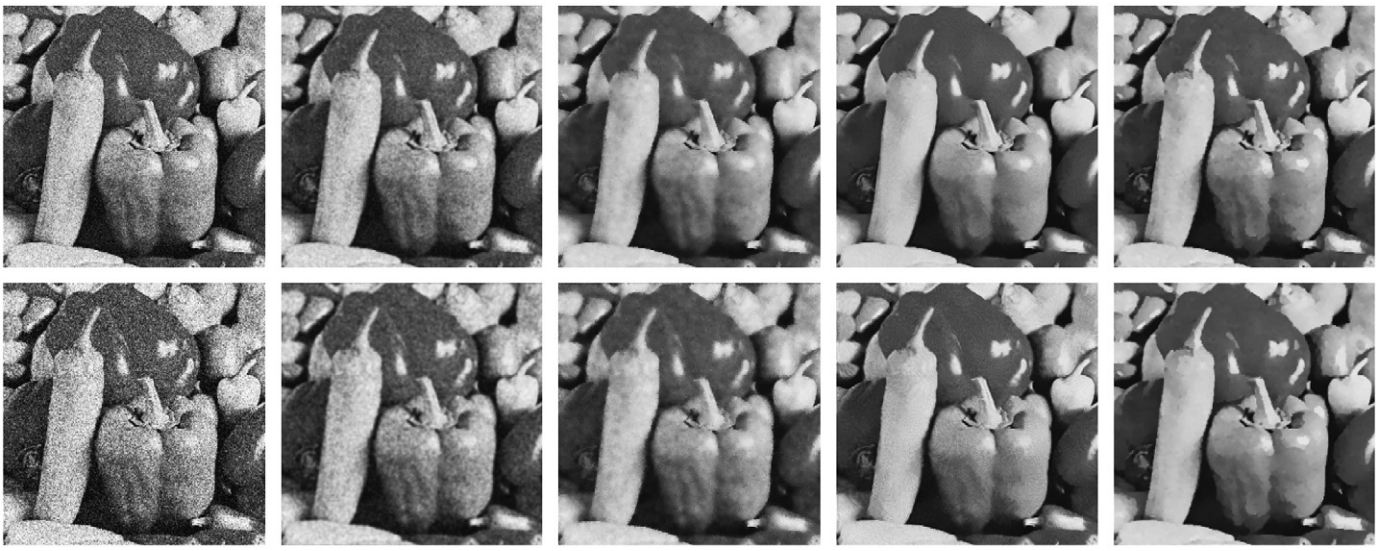

Fig. 6. Results of the state-of-art denoising algorithms on "Pepper" image with the noise $\sigma = 20$ (the first row) and $\sigma = 30$ (the second row). From left to right: the noisy image, the result of GF, BF, NL, and our algorithm.

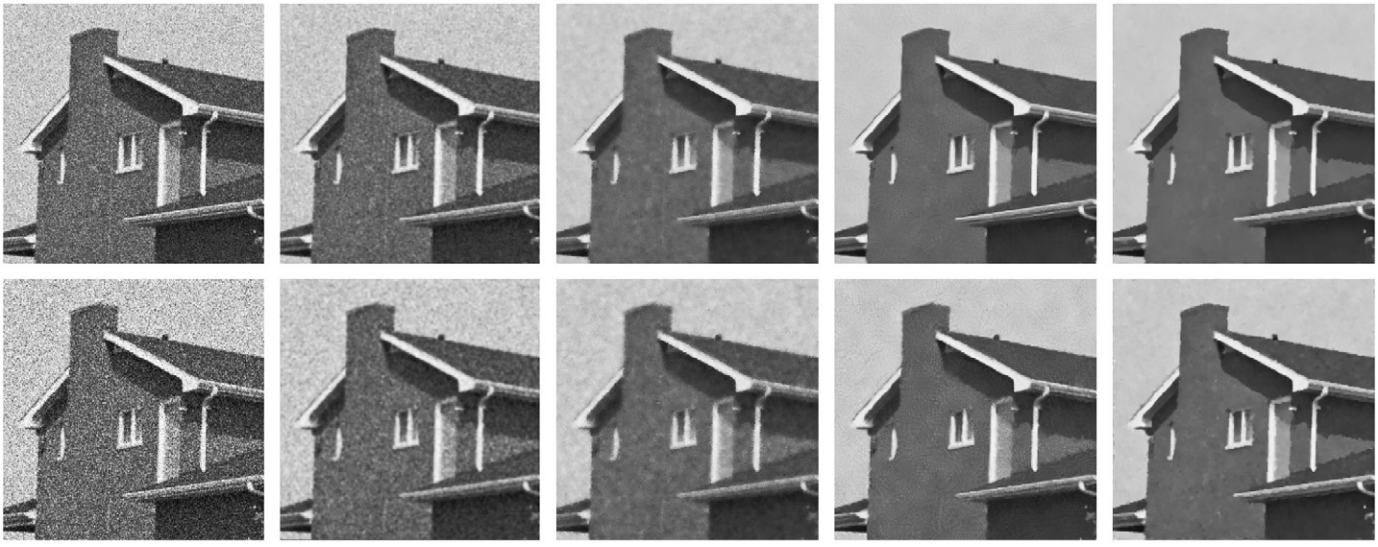

Fig. 7. Results of the state-of-art denoising algorithms on "House" image with the noise $\sigma = 20$ (the first row) and $\sigma = 30$ (the second row). From left to right: the noisy image, the result of GF, BF, NL, and our algorithm.

which are demonstrated to be efficient for denoising, optimal guarantee is attainable with Theorem 2. Table 2 displays the obtained continuous global optimal energy E_{opt} in our algorithm on five noisy images with $\sigma = 20$, the calculated discretization deviation bound using Theorem 2 denoted by *bound* and the proportion parameter with respect to E_{opt} represented by *factor*. We further show the final discretized results' energy. The data in the table strongly demonstrate the guaranteed optimal property of our algorithm and in practice the energy after discretization is much more closer to continuous optimum, which substantiates the optimization efficiency of our algorithm.

Furthermore, we compare the output energies of BP, GC and our algorithm on the same energy function to demonstrate the efficiency of our algorithm as an energy minimization algorithm. We choose quadratic data term and quadratic smooth term with constant weight $\lambda = 5$. Due to the semimetric priors, swap-GC is used for comparison together with max-product BP and our algorithm. Fig. 5 gives the intuitional energy comparisons on five noisy images ($\sigma = 20$). To make the comparisons vivid and further show the deviations from continuous global optimum, we plot the energy difference between outputs of each algorithm and globally optimum on the five images, respectively. It is clear that our algorithm achieves the lowest energy for all the five images and the energy difference is small, which further demonstrates the optimal property of our algorithms in practice. Further to note that max-product BP obtains the comparable energy with ours and swap-GC is much worse than BP and ours. The energy results are consistent with PSNR and MSE comparisons. Therefore, we can conclude that our algorithm outperforms BP and GC algorithms both in terms of denoising outputs and also the energy minimization outputs.

5.2. Comparison with the state-of-art algorithms

After the comparison with the MRFs based algorithms, we extend the experiments on the state-of-art spatial denoising algorithms. Gaussian Filter (GF) and Bilateral Filter (BF) are two classical denoising methods which are representative for their peers and NL-means (NL) is one of the most competitive denoising

method in spatial domain, to the best of our knowledge. Similarly, in the below comparisons, we also choose the best parameters for each algorithm, respectively, in terms of the best PSNR. According to the different noise level ($\sigma = 10$, 20, 30, 50 and 100), the filter window size ranges from 7 to 19 in GF and filter standard deviation is between 0.5 and 2 while filter window size, standard deviation of spatial domain and intensity domain are between 7 and 11, 1 and 2, 0.1 and 2, respectively, in BF. Meanwhile, the search window size in NL ranges from 15 to 21 and similarity window size is 5 or 7 with the filter degree between 100 and 1000. Further to note that the window size determines the algorithm running speed, especially for NL algorithm. The average running time for GF, BF and NL are about 0.5 s, 5 s, 7 min, respectively.

Figs. 6 and 7 exhibit the results on the noisy images "Pepper" and "House" with $\sigma = 20$ and 30. The figures indicate that the results of our algorithm and the NL algorithm look much better than others. The results of GF and BF are under-smooth and the blurring effect occurs. The NL and our algorithm can remove the noise and impose the consistency in the smooth regions effectively while preserving the details and edges well.

Table 3
PSNR and MSE comparison among the state-of-art denoising algorithms: Gaussian Filter(GF), Bilateral Filter(BF), Non-local Means Filter(NL) and Ours, on the noisy images "Pepper" and "House" with different level noise. $PSNR_0$ is the PSNR value of the noisy image.

$\sigma/PSNR_0$	PSNR				MSE			
	GF	BF	NL	Ours	GF	BF	NL	Ours
Pepper								
10/28.15	34.23	36.26	**36.79**	36.56	24.56	15.37	**13.62**	14.36
20/22.20	32.24	33.43	34.05	**34.24**	38.80	29.51	25.60	**24.51**
30/18.52	31.51	44.37	32.13	**32.78**	45.88	42.38	39.81	**34.25**
50/14.20	30.40	**30.87**	30.02	30.46	59.34	**53.23**	64.69	58.48
100/8.17	29.13	29.18	27.78	**29.21**	79.42	78.52	108.37	**77.99**
House								
10/28.11	34.81	37.22	**38.62**	37.55	21.46	12.32	**8.94**	11.42
20/22.14	32.81	34.29	**35.90**	35.76	34.06	23.66	**16.73**	17.25
30/18.58	32.09	32.60	33.27	**34.76**	40.22	35.74	30.66	**21.71**
50/14.19	31.01	31.60	30.72	**32.21**	51.51	45.03	55.03	**39.06**
100/8.09	30.17	30.06	27.90	**30.19**	62.46	64.16	105.65	**62.26**

The PSNR and MSE values are given in Table 3. From the results we can see that for less noisy images ($\sigma = 10$, 20), our algorithm obtains comparable PSNR and MSE values to the NL algorithm and is better than the other two algorithms. For more noisy images ($\sigma = 30$, 50, 100), our algorithm achieves the best results. It is worth noticing that the NL algorithm degrades sharply on strongly noisy images due

Table 4
PSNR and MSE values by the six algorithms on the "Lena" image. $PSNR_0$ is the PSNR value of the noisy image.

$\sigma/PSNR_0$		10/28.11	20/22.11	30/18.51	50/14.21	100/8.20
PSNR	GF	34.70	32.89	32.07	31.18	29.44
	BF	35.55	33.16	31.98	31.01	29.22
	NL	**36.29**	**33.98**	32.51	30.32	27.87
	BP	35.48	33.39	32.57	31.49	28.33
	GC	35.40	33.38	32.20	30.95	28.16
	Ours	35.77	33.65	**32.63**	**31.55**	**29.47**
MSE	GF	22.01	33.40	40.39	49.52	74.01
	BF	18.10	31.44	41.23	51.53	77.79
	NL	**15.29**	**26.03**	36.46	60.38	106.08
	BP	18.41	29.80	35.97	46.18	95.48
	GC	18.77	29.88	39.17	52.20	99.42
	Ours	17.21	28.04	**35.49**	**45.51**	**73.43**

Table 5
Average PSNR and MSE values on the 300 noisy images of the Berkeley segmentation data set. $PSNR_0$ is the PSNR value of the noisy image.

$\sigma/PSNR_0$		10/28.12	20/22.09	30/18.61	50/14.16	100/8.11
PSNR	GF	34.33	32.91	32.08	31.38	29.18
	BF	35.17	32.98	32.10	31.14	29.37
	NL	**35.71**	**33.64**	32.41	30.48	27.95
	BP	35.10	33.07	32.35	31.50	28.21
	GC	35.08	32.96	32.12	31.15	27.82
	Ours	35.59	33.60	**32.51**	**31.58**	**29.54**
MSE	GF	23.98	33.26	40.30	47.27	78.60
	BF	19.79	32.75	40.06	50.03	75.26
	NL	**17.48**	**28.13**	37.33	58.26	104.29
	BP	20.09	32.09	37.87	46.07	98.25
	GC	20.19	32.86	39.94	49.87	107.45
	Ours	17.96	28.36	**36.50**	**45.23**	**72.23**

Fig. 8. Results of the six algorithms on "Lena" image with the noise $\sigma = 20$ (the first row), $\sigma = 30$ (the second row) and $\sigma = 50$ (the third row). From left to right: the noisy image, the result by BP, GC, GF, BF, NL, and our algorithm.

to the corruption of its image structures similarity assumption, that is, a window in a natural image has many similar windows in the same image. In a strongly noisy image, the noise causes the structure similarity unconfident and lossy, therefore NL algorithm even generates many undesirable artificial textures in the smooth regions.

For the sake of better visual comparisons on all the algorithms (BP, GC, GF, BG, NL and Ours), we show the visual results of these algorithms on the "Lena" image with $\sigma = 20$, 30 and 50 in Fig. 8 and quantitative experimental results are given in Table. 4. The results further demonstrate the competitive even better performance of our algorithm.

Table 6
PSNR and MSE values by algorithm BM3D and the proposed algorithm on image "Lena", "Barbara", and "Boat" with the size of 512×512. $PSNR_0$ is the PSNR value of the noisy image.

σ/$PSNR_0$		10/28.14	20/22.11	30/18.59	50/14.16	100/8.14
Lena						
PSNR	BM3D	**35.93**	33.05	31.26	29.05	25.95
	Ours	35.67	**34.36**	**32.03**	**31.79**	**29.32**
MSE	BM3D	**16.60**	32.22	48.65	80.92	165.23
	Ours	17.62	**23.84**	**40.78**	**43.02**	**76.04**
Barbara						
PSNR	BM3D	**34.98**	31.78	29.81	27.23	23.62
	Ours	34.07	**32.25**	**31.52**	**30.79**	**28.45**
MSE	BM3D	**20.66**	43.16	67.93	123.05	282.54
	Ours	25.45	**38.77**	**45.82**	**54.15**	**93.00**
Boat						
PSNR	BM3D	33.92	30.88	29.12	26.78	23.97
	Ours	**34.95**	**32.66**	**31.71**	**31.17**	**28.05**
MSE	BM3D	26.37	53.10	79.63	136.48	260.66
	Ours	**20.82**	**35.26**	**43.86**	**49.67**	**101.87**

Another experiment is carried out on the Berkeley segmentation benchmark [31]. 300 nature images are also corrupted by Gaussian noise with $\sigma = 10$, 20, 30, 50, and 100. The parameters for each algorithm are the same as those obtained in the first experiment. Table 5 shows the average PSNR and MSE values. From this table, it is obvious that our algorithm performs the best.

At last, we test our algorithm and one of the best state-of-the-art method, the BM3D [18], on image "Lena", "Barbara", and "Boat" with the size of 512×512. The comparative results are given in Table 6. The results of the BM3D are copied from the authors' webpage http://www.cs.tut.fi/foi/GCF-BM3D/. From the result we can see that our algorithm has comparable results to the BM3D for images with less noise; while performs better than the BM3D for more noisy images. The running time of our algorithm on a 512×512 image is about 20 s.

The results of visual comparisons of BM3D and our algorithm are given in Fig. 9. In this figure, we give the results of our algorithm with $a=1$, $a=2$, and $a=3$ in the right three columns, respectively. From the results we can see that our results tend to smoothness compared to BM3D. The different setting of parameter a affects the level of smoothness. The smoothness-tending results of our algorithm may be not optimal for human perception. However, the best numerical results (PSNR and MSE values) indicate that our algorithm is more suitable for preprocessing in applications of image processing and computer vision.

6. Conclusion

In this paper, a novel image denoising algorithm has been proposed. This algorithm formulates the image denoising problem as an energy minimization problem based on an MAP-MRF

Fig. 9. Results of BM3D and our algorithm on 512×512 "Lena" image with the noise $\sigma = 20$ (the first row), $\sigma = 30$ (the second row) and $\sigma = 50$ (the third row). From left to right: the noisy image, the result by BM3D, our algorithm with $a=1$, $a=2$, and $a=3$.

model. Since such an optimization problem is NP-hard in discrete sense, there is no guarantee for an optimization algorithm to obtain an exact global optimum and many optimization techniques denoise the image iteratively and therefore cannot have the closed form solution. In our algorithm, by relaxing the image intensities from discrete values to continuous values, a closed-form globally optimal solution is obtained. Although the deviation from continuous globally optimum is inevitable since the discretization must be carried on to present the digital image, we can obtain an upper bound for the deviation theoretically which demonstrates the optimal property of our algorithm. Moreover, the pre-estimated edge information is incorporated in our formulation to preserve edges well during denoising. Meanwhile, the utilization of the Laplacian matrix based on the weighted graph construction commendably quickens up our algorithm which also facilitates our closed form solution formulation. In comparison with model variants, our algorithm outperforms the other two mature algorithms both on the image denoising task and on the energy minimization. The extensive comparative experiments with state-of-the-art methods demonstrate that our proposed approach performs competitively, and often better than the best existing ones.

As the same as many existing algorithms, the proposed algorithm meets the problem of parameter selection (parameters a and b in (8)). In our experiments, a is selected within a reasonable range and b is fixed as the same value. One possible solution to this problem is to learn the parameters via maximizing marginal probability from a set of training data.

In addition, as a good solving technique for MRFs energy optimization, our algorithm can be efficiently applied to many other tasks, e.g. image segmentation, stereo and motion and image synthesis in image processing and computer vision, which can be modeled as MAP-MRFs. The formulation and optimal property analysis can be followed in a similar way.

Acknowledgments

This work was supported by the Introduced Innovative R&D Team of Guangdong Province (Robot and Intelligent Information Technology); Natural Science Foundation of China (60903117); and Science, Industry, Trade, and Information Technology Commission of Shenzhen Municipality, China (JC201005270357A, ZYC201006130314A). We are thankful also to the anonymous reviewers for their constructive suggestions which helped us improving our manuscript.

References

[1] D. Donoho, De-noising by soft-thresholding, IEEE Transactions on Information Theory 41 (1995) 613–627.
[2] F. Luisier, T. Blu, M. Unser, Image denoising in mixed Poisson–Gaussian noise, IEEE Transactions on Image Processing 19 (2011) 696–708.
[3] M. Lindenbaum, M. Fischer, A. Bruckstein, On Gabor's contribution to image enhancement, Pattern Recognition 27 (1994) 1–8.
[4] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259–268.
[5] M. Lysaker, X. Tai, Iterative image restoration combining total variation minimization and a second-order functional, International Journal of Computer Vision 66 (2006) 5–18.
[6] E. Tadmor, S. Nezzar, L. Vese, A multiscale image representation using hierarchical (bv,l[sup 2]) decompositions, Multiscale Modeling and Simulation 2 (2004) 554–579.
[7] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Modeling and Simulation 4 (2005) 460–489.
[8] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 629–639.
[9] F. Catte, P. Lions, J. Morel, T. Coll, Image selective smoothing and edge detection by nonlinear diffusion, SIAM Journal on Numerical Analysis 29 (1992) 182–193.
[10] S. Smith, J. Brady, Susan‡a new approach to low level image processing, International Journal of Computer Vision 23 (1997) 45–78.
[11] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Proceedings of International Conference on Computer Vision, 1998, p. 839.
[12] M. Mignotte, Image denoising by averaging of piecewise constant simulations of image partitions, IEEE Transactions on Image Processing 16 (2007) 523.
[13] M. Mignotte, A segmentation-based regularization term for image deconvolution, IEEE Transactions on Image Processing 15 (2006) 1973.
[14] M. Mignotte, An adaptive segmentation-based regularization term for image restoration, in: Proceedings of IEEE International Conference on Image Processing, vol. 1, 2005.
[15] M. Hansen, W. Higgins, Watershed-based maximum-homogeneity filtering, IEEE Transactions on Image Processing 8 (1999) 982–988.
[16] A. Buades, B. Coll, J. Morel, A non-local algorithm for image denoising, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2005, pp. 60–65.
[17] A. Buades, B. Coll, J. Morel, Nonlocal image and movie denoising, International Journal of Computer Vision 76 (2008) 123–139.
[18] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, IEEE Transactions on Image Processing 16 (2007) 2080–2095.
[19] P. Bouboulis, K. Slavakis, S. Theodoridis, Adaptive kernel-based image denoising employing semi-parametric regularization, IEEE Transactions on Image Processing 19 (2010) 1465–1479.
[20] S. Geman, D. Geman, K. Abend, T. Harley, L. Kanal, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, Journal of Applied Statistics 20 (1993) 25–62.
[21] J. Besag, On the statistical analysis of dirty pictures, Journal of the Royal Statistical Society. Series B (Methodological) 48 (1986) 259–302.
[22] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 1222–1239.
[23] V. Kolmogorov, R. Zabih, What energy functions can be minimized via graph cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004) 147–159.
[24] 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 (2004) 1124–1137.
[25] Y. Weiss, W. Freeman, On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs, IEEE Transactions on Information Theory 47 (2001) 736–744.
[26] P. Felzenszwalb, D. Huttenlocher, Efficient belief propagation for early vision, International Journal of Computer Vision 70 (2006) 41–54.
[27] M. Liu, S. Chen, J. Liu, Continuous mrf based image denoising with a closed form solution, in: IEEE International Conference on Image Processing, 2010, pp. 1137–1140.
[28] Y. Boykov, O. Veksler, R. Zabih, Markov random fields with efficient approximations, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1998, pp. 648–655.
[29] P. Jodoin, M. Mignotte, An energy-based framework using global spatial constraints for the stereo correspondence problem, in: Proceedings of International Conference on Image Processing, vol. V, 2004, pp. 3001–3004.
[30] S. Chao, D. Tsai, An improved anisotropic diffusion model for detail- and edge-preserving smoothing, Pattern Recognition Letters 31 (2010) 2012–2023.
[31] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of International Conference on Computer Vision, vol. 2, 2001, pp. 416–423.

Shifeng Chen received the B.E. degree from the University of Science and Technology of China, Hefei, in 2002, the M.Phil. degree from City University of Hong Kong, Hong Kong, China, in 2005, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, China, in 2008. He is now an associate professor in the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. His research interests include image/video processing and computer vision.
Ming Liu received the B.E. degree from the University of Science and Technology of China, Hefei, in 2006, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, China, in 2011. His research interests include image processing and pattern recognition.

Wei Zhang received the B.S. degree from Nankai University, China, in 2002, the M.E. degree from Tsinghua University, China, in 2005, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, China, in 2010. He is now a research assistant professor in the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. His research interests include pattern recognition, image processing and computer vision.

Jianzhuang Liu (M'02–SM'02) received the B.E. degree from the Nanjing University of Posts and Telecommunications, PR China, in 1983, the M.E. degree from the Beijing University of Posts and Telecommunications, PR China, in 1987, and the Ph.D. degree from The Chinese University of Hong Kong, Hong Kong, in 1997. From 1987 to 1994, he was a faculty member in the Department of Electronic Engineering, Xidian University, PR China. From August 1998 to August 2000, he was a research fellow at the School of Mechanical and Production Engineering, Nanyang Technological University, Singapore. Then he was a postdoctoral fellow and an assistant professor in The Chinese University of Hong Kong for several years. He is now a research professor in the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. His research interests include computer vision, image processing, machine learning, and graphics.