

A Framework for Moving Least Squares Method with Total Variation Minimizing Regularization

Yeon Ju Lee · Sukho Lee · Jungho Yoon

Published online: 6 March 2013
© Springer Science+Business Media New York 2013

Abstract In this paper, we propose a computational framework to incorporate regularization terms used in regularity based variational methods into least squares based methods. In the regularity based variational approach, the image is a result of the competition between the fidelity term and a regularity term, while in the least squares based approach the image is computed as a minimizer to a constrained least squares problem. The total variation minimizing denoising scheme is an exemplary scheme of the former approach with the total variation term as the regularity term, while the moving least squares method is an exemplary scheme of the latter approach. Both approaches have appeared in the literature of image processing independently. By putting schemes from both approaches into a single framework, the resulting scheme benefits from the advantageous properties of both parties. As an example, in this paper, we propose a new denoising scheme, where the total variation minimizing term is adopted by the moving least squares method. The proposed scheme is based on splitting methods, since they make it possible to express the minimization problem as a linear system. In this paper, we employed the split Bregman scheme for its simplicity. The resulting denoising scheme overcomes

the drawbacks of both schemes, i.e., the staircase artifact in the total variation minimizing based denoising and the noisy artifact in the moving least squares based denoising method. The proposed computational framework can be utilized to put various combinations of both approaches with different properties together.

Keywords Denoising · Total variation · Moving least squares · Bregman iteration

1 Introduction

In the last two decades, the regularity based variational approach has aroused a large impact on the image processing society [1, 17, 46]. Most of the models in this approach take the form of an unconstrained regularized data fitting model, where the desired image is obtained as a regularized minimizer to a certain energy functional which contains both fidelity and regularization terms. For example, the total variation based image restoration model [14, 45], which has become one of the most popular image denoising models due to its edge preserving property, can be described as an unconstrained problem with the total variation as the regularization term.

Often, it is hard, if not impossible, to find a regularized minimizer solution in closed form for the regularized variational model. Therefore, iterative schemes are exploited to obtain the regularized minimizer. In earlier works, iteration schemes have been written in the Euler-Lagrange type, which often result in nonlinear *partial differential equations* (PDEs), which are normally slow and not applicable for industrial purposes [1, 55]. To accelerate the computation, several real-time working discretizations of Euler-Lagrange type equations have been proposed. In [9], a real-time working discretization of the Euler-Lagrange equation is proposed for the estimation of the optical flow field, which

Y.J. Lee (✉)
Institute of Mathematical Sciences, Ewha W. Univ.,
Seoul 120-750, South Korea
e-mail: leeyeonju08@ewha.ac.kr

S. Lee
Dept. Software Engineering, Dongseo Univ., Busan 617-716,
South Korea
e-mail: petrasuk@gmail.com

J. Yoon
Department of Mathematics, Ewha W. Univ., Seoul 120-750,
South Korea
e-mail: yoon@ewha.ac.kr

combines two hierarchical strategies. Bidirectional multi-grid methods have also been applied to accelerate the speed of a broad class of variational optic flow models [10]. Furthermore, specific algorithms designed for parallel hardware make real-time computation possible [31].

Recently, the iterative schemes have undergone a lot of optimization, and faster iteration schemes, such as Bregman-type iteration schemes, have been developed [2, 30, 43, 57, 61]. The development of iterative schemes not only has resulted in faster computation but also has made it easier to incorporate nonlinear regularization terms into variational functionals.

Meanwhile, another type of variational approach, which is in its nature a nonparametric data fitting approach and is based on localized least squares method, has recently appeared in the literature of image processing [5, 20, 27, 52, 60]. In this approach, the problem is formulated in the form of a constrained data fitting model. The characteristics of this approach is that it takes the data and its position to define the structure of the approximation model, and therefore, the local characteristics of the data can be taken into account in the approximation. One of the most popular method in this framework is the *moving least squares* (MLS) model [35]. The MLS model constructs and evaluates a local polynomial fit continuously over the entire domain, resulting in the MLS fit function. The MLS model has shown good performance for interpolation based image processing such as image zooming and superresolution [5, 20, 27, 52, 60]. Normally, the MLS model seeks to find a solution in closed form, which is often the case with methods in the least squares framework.

In this paper, we propose to put the MLS model into the iterative regularity based variational framework. By putting the MLS model in the iterative framework, it becomes possible for the MLS model to adopt several well-known regularization terms that were formerly proposed in other regularized variational problems. To illustrate an example, we show in this paper that the well-known *total variation* (TV) minimizing regularizer can be successfully adopted in the MLS model in the proposed framework. We want to emphasize the fact that, until now, there exists no method which has successfully adopted the TV regularizer into the MLS model using the conventional least squares framework.

The adoption of the TV regularization term endows the MLS model with better denoising property. Normally, MLS based algorithms are weak against noise, since, in general, least squares methods are weak against outliers. In contrast, TV based methods are strong against outliers, since outliers have large variation values. The TV regularization term eliminates the outliers very fast and helps the regularized MLS to produce a better approximation to the original noise-free image.

Furthermore, the proposed scheme can also provide a better solution than conventional TV schemes. Conventional

TV schemes process the observed image towards a piecewise constant image, which exhibits many false jump discontinuities and is visually unpleasant. This phenomenon is called the *staircase effect*. The staircase effect of the TV denoising is due to the fact that the total variation is not differentiable and the differential equations involved in this problem are at most second order. This makes the solution of the TV denoising to have a first order approximation. By incorporating the constraints used in the MLS problem into the proposed scheme, the solution becomes more constrained achieving a higher order approximation than that of the conventional TV denoising. As a result, smooth regions are well preserved, and the staircase effect is avoided without sacrificing the edge preserving property too much, as is the case with schemes trying to obtain a solution using higher order regularization terms [16, 18, 19, 51, 56].

A main difference between the proposed model and other regularized variational models is that we obtain an energy functional, whose minimization is done with respect to the parameters used in the reconstruction of the image rather than the image itself. In other words, instead of working on the image, we are working on the localized parameters used to reconstruct a local region of the image, which gives us the chance of finer control on the solution image. The proposed framework can benefit from extensions of both the regularized variational framework and the MLS framework. It is possible to generate combinations of both extensions, resulting in numerous denoising schemes with different properties.

2 Previous Works

Several different algorithms are used in the construction of the proposed scheme. In this section, we briefly review the previous works related to the different algorithms used.

The *Moving Least Squares* (MLS) approximation was introduced first in [34], and the details on the approximation power, stability, and implementational issues related to MLS surfaces are presented in [35]. Since then, the MLS method has been applied to many areas. For example, a diffuse approximation method is proposed in [40] to replace the FEM interpolation for solving partial differential equations. The MLS is generalized in [38], and it is proved that it produces the diffuse derivatives introduced in [40]. In [7], the MLS approximation applied to optimization problems, such as the construction of response formulation and the optimization of sheet forming method. The MLS technique is also related to local M-smoothers [20, 58] in the aspect that it approximates the local data by certain approximation functions. The usage of such approximation in variational/regularization method has been the topic of [39] and [44].

Recently, the MLS method has been proved to be quite useful in interpolation based image processing such as super-resolution and image zooming [5, 20, 27, 52, 60]. The work in [27] applies MLS for image denoising by using two different kinds of robust estimators. A paper that uses a KR-based method as an edge-preserving smoother is given in [20]. In [52], a steering kernel regression is proposed which steers the local kernels along the directions of the local edge structure. Bose and Ahuja applied the MLS to noise filtering and approximation of irregularly spaced data in the super-resolution problem [5]. Yoon and Lee employed the Gaussian radial basis functions instead of algebraic polynomials for an edge-directed image upsampling method [60]. However, the models show weakness against outliers, which is the main reason that the denoising properties of MLS based models normally are not good.

Meanwhile, the total variation based denoising model was first introduced in [45] in the following form:

$$\min_u \|\nabla u\|_1 + \frac{\mu}{2} \|u - f\|_2^2, \quad (1)$$

where u is the solution image, f is the measured noisy image, and ∇ is the gradient operator. The term $\|\nabla u\|_1$ is called the total variation norm, and μ controls the balance between the fidelity term and the total variation term. The minimization of the total variation term smooths out very fast small-scaled variations which results in removal of the noise. Compared to the minimizing of the L_2 norm of the image gradient, the minimizing of the total variation has the property of preserving sharp discontinuities (edges) in images while removing the noise. This is a desirable property for images, since the visual quality of an image greatly depends on the preservation of edges.

However, the minimizing of the TV norm also makes smooth regions flat which makes the image look unnatural. This is known as the staircasing effect [4, 42]. As shown in [12] and [41], the staircasing effect of the total variation based denoising is due to the fact that the total variation is not differentiable and the total variation minimization tries to approximate the image to an approximation of first order. Therefore, to obtain higher order approximation of the solution image, smooth approximations and variants of the total variation functional have been proposed to avoid the staircasing effect [4, 16, 18]. In [14], the infimal convolution of functionals with first and second order derivatives as regularizers is used to obtain a higher order approximation. General modified infimal convolution functionals which combine L_1 norms with linear operators and which can be implemented with fast algorithms are used in [50].

Extra constraint terms have also been proposed to preserve the smooth areas. For example, in [56], they introduce a constraint term for smooth area protection as a penalty function. In [19], a second-order derivative is introduced

into the functional to reduce the penalty for slow variation in smooth regions. These approaches try to obtain a higher order approximation by modifying the regularization function or the fidelity function. However, by modifying the regularization function, the nice geometrical properties of the total variation and the fast and strong denoising property may be lost.

Another important topic relating the total variation minimization is the implementation issue. At the early stage, Euler-Lagrange formulations have been used to implement the total variation minimization. However, the Euler-Lagrange formulation contains the non-differentiable term $\nabla \cdot \frac{\nabla u}{|\nabla u|}$ which can not be computed exactly at $\nabla u = 0$. Therefore, splitting methods have been exploited. Using operator splitting methods, the term $\nabla \cdot \frac{\nabla u}{|\nabla u|}$ can be split into sub-operators of simpler forms which can be implemented easier. Osher proposed the split Bregman algorithm to solve the TV problem using the Bregman iteration [30]. The equivalence of the split Bregman algorithm and the augmented Lagrangian method have been shown in [25, 49, 59]. In [48], it is shown that the alternating split Bregman algorithm can be interpreted also as a Douglas-Rachford splitting algorithm [36] which is known to be a special case of the proximal point algorithm [23]. Meanwhile, Chambolle proposed an algorithm based on a dual formulation which is fast and effective for minimizing the total variation [13]. When the dual part becomes more complicated primal-dual formulations can be used, which update both a primal and a dual variable. Zhu and Chan proposed an efficient primal-dual hybrid gradient algorithm for total variation image restoration [63]. Esser et al. proposed a modified version of the primal-dual hybrid gradient algorithm proposed by Zhu and Chan [26]. Zhang et al. proposed a unified primal-dual algorithm framework for two classes of problems [62]. Currently, one of the most efficient primal-dual splitting schemes is the scheme of Chambolle and Pock which shows a fast convergence for problems where the primal or the dual objective is uniformly convex [15]. As will be explained in Sect. 3, we exploit splitting methods to get an explicit form for the sought solution.

In the next subsections, we give a detailed explanation of the used methods in our formulation.

2.1 Moving Least Squares Based Denoising

The solution of the MLS method has a closed form and is easily computed by solving a linear system. To explain the MLS method, we use the following notations: Suppose that our observed image is a discrete sampling of a function $f : \mathcal{D} \rightarrow \mathbb{R}$ at an equally spaced point set in a rectangle $\mathcal{D} \subset \mathbb{R}^2$. We want to find an overall MLS approximation function Lf of the data function f . For this aim, we first construct local approximation functions $p := p_x$ for each

point $\mathbf{x} \in \mathcal{D}$. Then the overall MLS approximation function Lf is obtained by collecting all the function values $p_{\mathbf{x}}(\mathbf{x})$ from each local approximation function $p_{\mathbf{x}}$ at the points $\mathbf{x} \in \mathcal{D}$:

$$Lf(\mathbf{x}) = p_{\mathbf{x}}(\mathbf{x}), \quad \forall \mathbf{x}. \tag{2}$$

For example, let us derive the local approximation functions from a polynomial space of degree L , denoted by Π_L , where the degree of a polynomial space represents the highest degree of a polynomial in the polynomial space. We denote by m the dimension of Π_L , i.e., the number of basis in Π_L , which is calculated as

$$m := \dim \Pi_L = \frac{(L+2)(L+1)}{2},$$

and denote by $\{p_\ell\}_{\ell=1}^m$ the basis of Π_L . For example, for $L = 2$, $m = \frac{(2+2)(2+1)}{2} = 6$, we get the polynomial space Π_2 with the corresponding basis $\{p_\ell\}_{\ell=1}^6 = \{1, x, y, x^2, xy, y^2\}$. To solve a local polynomial approximation function p around a certain point \mathbf{x} , a finite set of data points around \mathbf{x} have to be involved into the computation. Let $\mathcal{N}(\mathbf{x})$ be a finite set of distinct data points in \mathbb{R}^2 around \mathbf{x} , and N be the number of points in $\mathcal{N}(\mathbf{x})$. Usually, N has to be bigger than m so that the problem can be solved. Then, for a point $\mathbf{x} \in \mathcal{D}$, the polynomial approximation function p is obtained by solving the following quadratic minimization problem:

$$\operatorname{argmin}_{p \in \Pi_L} \left\{ \sum_{n=1}^N \|p(\mathbf{x}_n) - f(\mathbf{x}_n)\|_2^2 \theta(\|\mathbf{x} - \mathbf{x}_n\|_2) \right\}, \tag{3}$$

where θ is a radial and fast decreasing smooth function and $N > \dim \Pi_L$. The θ function is a weight function which decides how much the data point $f(\mathbf{x}_n)$ contributes to the construction of the polynomial approximation $p(\mathbf{x}_n)$. If $\theta(\|\mathbf{x} - \mathbf{x}_n\|_2)$ is large, $f(\mathbf{x}_n)$ contributes much to the construction. This is done by reducing the error of the data $f(\mathbf{x}_n)$ and the evaluated value $p(\mathbf{x}_n)$ at \mathbf{x}_n , i.e., by reducing $\|p(\mathbf{x}_n) - f(\mathbf{x}_n)\|_2^2$. A typical choice of θ is the Gaussian function

$$\theta(\mathbf{r}) = e^{-|\mathbf{r}|^2/\sigma} \tag{4}$$

having a scale parameter $\sigma > 0$. This is based on the premise that the data points near to \mathbf{x} should have a large effect in the construction. The minimizer p differs from pixel to pixel. Since we construct p in Π_L of dimension m , we can write p as a linear combination of the basis $\{p_\ell\}_{\ell=1}^m$:

$$p := \sum_{\ell=1}^m c_\ell p_\ell. \tag{5}$$

Therefore, we see that (3) is a least squares problem that can be solved for the coefficients $\{c_\ell : \ell = 1, \dots, m\}$ of the

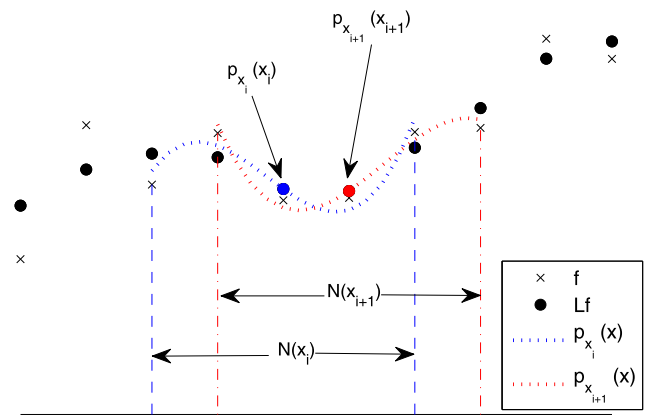


Fig. 1 Exemplary illustration showing the construction of the MLS approximation for a 1-D data profile

polynomial p . The coefficients $\{c_\ell : \ell = 1, \dots, m\}$ satisfying (3) are found by calculating the derivative of the function in (3) with respect to $\{c_\ell : \ell = 1, \dots, m\}$ and setting them equal to zero:

$$\frac{\partial (\sum_{n=1}^N \|p(\mathbf{x}_n) - f(\mathbf{x}_n)\|_2^2 \theta(\|\mathbf{x} - \mathbf{x}_n\|_2))}{\partial c_\ell} = 0, \tag{6}$$

$$\ell = 1, \dots, m.$$

When the solution, say $p_{\mathbf{x}}(\mathbf{x})$, of the above quadratic minimization problem is obtained for a certain point $\mathbf{x} \in \mathcal{D}$, then the function value $Lf(\mathbf{x})$ for the point \mathbf{x} becomes $Lf(\mathbf{x}) = p_{\mathbf{x}}(\mathbf{x})$. Thus, the overall MLS approximation function Lf is defined as follows:

$$Lf(\mathbf{x}) = p_{\mathbf{x}}(\mathbf{x}), \quad \forall \mathbf{x}, \tag{7}$$

where $p_{\mathbf{x}}(\cdot)$ denotes the minimizer of (3) for the point \mathbf{x} . Figure 1 is an exemplary illustration showing the concept of constructing the MLS approximation function for a 1-D data profile. The blue and the red dotted lines show the constructed polynomial functions $p_{\mathbf{x}_i}$ and $p_{\mathbf{x}_{i+1}}$ for the points \mathbf{x}_i and \mathbf{x}_{i+1} , respectively, using the data in the neighborhoods $\mathcal{N}(\mathbf{x}_i)$ and $\mathcal{N}(\mathbf{x}_{i+1})$, respectively. The value $p_{\mathbf{x}_i}(\mathbf{x}_i)$ denotes the value of $p_{\mathbf{x}_i}$ at the point \mathbf{x}_i , and likewise for $p_{\mathbf{x}_{i+1}}(\mathbf{x}_{i+1})$. The overall approximation MLS function Lf becomes the collection of all the values $p_{\mathbf{x}}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{D}$, which are shown as black circle points in Fig. 1.

If f in (3) is a noisy version of the original data, the approximation Lf becomes a smoother one. By obtaining a smoother approximation Lf , denoising is achieved. However, it is well-known that the MLS method has limitations in denoising, because the output of least squares method is highly affected by singular data (or outlier) error, which results in noisy artifacts. The proposed scheme aims to reduce this artifact by introducing the TV norm into the least squares method.

2.2 Split Bregman Method

Even though other splitting methods can be applied to the proposed scheme, as will be explained in Sect. 3, we exploit the split Bregman method for its simplicity. The split Bregman method [30, 61] solves a minimization problem by operator splitting and then applying Bregman iteration to solve the split problem. It solves the problem

$$\min_u \|\Phi(u)\|_1 + E(u)$$

by the operator splitting

$$\min_{u,d} \|d\|_1 + E(u) \quad \text{subject to} \quad \Phi(u) = d,$$

where E is convex and Φ is convex and differentiable. For image processing applications, u is the solution image, and d is used to obtain the constraint formulation. Applying Bregman iteration with $J(u, d) = \|d\|_1 + E(u)$ and $H(u, d) = \frac{1}{2}\|d - \Phi(u)\|_2^2$ yields

$$\begin{aligned} (u^{k+1}, d^{k+1}) = \min_{u,d} J(u, d) - \langle p_u^k, u - u^k \rangle - \langle p_d^k, d - d^k \rangle \\ + \lambda H(u, d) \end{aligned} \tag{8}$$

$$p_u^{k+1} = p_u^k - \lambda \nabla_u H(u^{k+1}, d^{k+1})$$

$$p_d^{k+1} = p_d^k - \lambda \nabla_d H(u^{k+1}, d^{k+1})$$

where p_u and p_d are in the set of the sub-gradients of E with respect to u and d , respectively, λ is a parameter that controls the balance, k and $k + 1$ denote the current and the next iteration step, and $\nabla_u H(u, d) = (\nabla \Phi(u))^*(\Phi(u) - d)$ and $\nabla_d H(u, d) = d - \Phi(u)$.

It is shown in [30] that (8) is equivalent to the following two subproblems:

$$\begin{aligned} (u^{k+1}, d^{k+1}) = \min_{u,d} \|u\|_1 + E(u) + \frac{\lambda}{2} \|d - \Phi(u) - b^k\|_2^2 \\ b^{k+1} = b^k + (\Phi(u^{k+1}) - d^{k+1}). \end{aligned} \tag{9}$$

Here, the first subproblem works on the primal variable u , on the image itself, and the second one works on the dual variable b , which can be seen as the noise. The first subproblem can be solved by alternately minimizing u and then d ,

$$\begin{aligned} u^{k+1} = \arg \min_u E(u) + \frac{\lambda}{2} \|d^k - \Phi(u) - b^k\|_2^2 \\ d^{k+1} = \arg \min_d \|d\|_1 + \frac{\lambda}{2} \|d - \Phi(u^{k+1}) - b^k\|_2^2. \end{aligned} \tag{10}$$

The step for u^{k+1} can be solved as a linear system, while in the step for d^{k+1} , the minimizer can be expressed in closed-form as a shrinkage. For the total variation problem, $E(u)$

and $\Phi(u)$ becomes $E(u) = \|u - f\|_2^2$ and $\Phi(u) = \nabla u$, and thus the problem is solved by the following iterative 3 steps:

$$\begin{aligned} (1) \quad u^{k+1} &= \min_u \frac{\mu}{2} \|u - f\|_2^2 + \frac{\lambda}{2} \|d^k - \nabla u - b^k\|_2^2 \\ (2) \quad d^{k+1} &= \min_d \|d\|_1 + \frac{\lambda}{2} \|d - \nabla u^{k+1} - b^k\|_2^2 \\ &= \text{shrink}(\nabla u^{k+1} + b^k, 1/\lambda) \\ (3) \quad b^{k+1} &= b^k + \nabla u^{k+1} - d^{k+1}. \end{aligned} \tag{11}$$

Here, the shrinkage operator is defined as follows:

$$\text{shrink}(x, \tau) = \begin{cases} x - \tau \text{sgn}(x) & \text{if } |x| \geq \tau, \\ 0 & \text{if } |x| < \tau, \end{cases}$$

where $\text{sgn}(\cdot)$ is the sign operator.

By using the split Bregman method, the total variation problem can be solved faster than using conventional non-linear partial differential equations. However, in this paper, we do not use the Bregman method for fast computation, but utilize and vary it to introduce the TV norm into the MLS model.

3 Proposed Scheme

In this section, we first propose a new energy functional to be minimized for data approximation. Then, we propose a new computational framework that can solve the minimization problem. Let f be a given reference image defined on a domain Ω , and let \mathbf{x} be a fixed point in Ω . Then, we construct a local polynomial approximation of degree L :

$$p(\mathbf{r}) := p_{\mathbf{x}}(\mathbf{r}) := \sum_{|\alpha|_1 \leq L} c_{\alpha} \mathbf{r}^{\alpha}$$

by minimizing the following energy functional:

$$\arg \min_{p \in \Pi_m} \left\{ \sum_{n=1}^N \|\nabla p(\mathbf{x}_n)\|_1 + \frac{\mu}{2} \|p(\mathbf{x}_n) - f(\mathbf{x}_n)\|_2^2 \theta(\mathbf{x}, \mathbf{x}_n) \right\}. \tag{12}$$

In this construction, we use a nonlocal weight function θ which is defined as

$$\theta(\mathbf{x}, \mathbf{r}) = \exp \left\{ - \frac{G_a * (\|f(\mathbf{x} + \cdot) - f(\mathbf{r} + \cdot)\|_2^2)(0)}{h_0^2} \right\},$$

where h_0 is a small positive value, G_a is the Gaussian function with standard deviation a , and $+\cdot$ denotes the shifting operator. Nonlocal similarity measures have been used in nonlocal denoising methods to decide whether a certain pixel should be used in the denoising process or

not [6, 8, 21, 22, 24, 28, 29, 33, 44, 53]. In comparison, we use it as a weight function such that the weight function becomes data adaptive and consider the similarity of the local areas between two positions \mathbf{x} and \mathbf{r} . The weight function has a large value if the local areas around \mathbf{x} and \mathbf{r} are similar. In this case, the $f(\mathbf{r})$ value is taken much into account in constructing the local approximation at \mathbf{x} . This is based on the premise that the approximation function value should be similar to the data value if the corresponding neighborhoods are similar. Thus, pixel values which are spatially distant are also incorporated into the approximation process if the corresponding θ value is large.

We construct p at each pixel in the image, and so we are solving the minimization problem in (12) as many times as there are pixels in Ω . Then, the overall approximation function Lf becomes $Lf(\mathbf{x}) := p(\mathbf{x}) := p_{\mathbf{x}}(\mathbf{x})$ for all $\mathbf{x} \in \Omega$.

Without the first term, the energy functional in (12) becomes just that used in the conventional MLS model. The incorporation of the TV norm endows the MLS model with several desirable properties that helps to overcome the weaknesses of the MLS model. Again, we want to emphasize the fact that even though the incorporation of the TV term may seem simple, the problem cannot be solved in the conventional moving least squares framework, and therefore, has never been proposed. This is due to the fact that a solution of the problem (12) having a closed form cannot be obtained. This is in contrast with TV approaches, for which quadratic data penalization have been frequently used with the TV norm [3, 32, 55].

The proposed model is also different from the conventional TV approach, in the sense that the solution sought

is not the image u , but are the parameters that are used in the reconstruction of the p function which approximates the image from the polynomial space. We call the proposed method the *MLS with TV minimizing* (MLS-TV) model. We will first explain the properties of the MLS-TV model by explaining its contributions to the conventional MLS and the TV models. Then, we will propose a computational framework which solves the problem in (12). The proposed framework can be applied for denoising schemes with other regularization terms.

3.1 Contribution of the MLS-TV Model to the MLS Model

By incorporating the TV norm, the proposed model gets the following advantageous properties over the conventional MLS model:

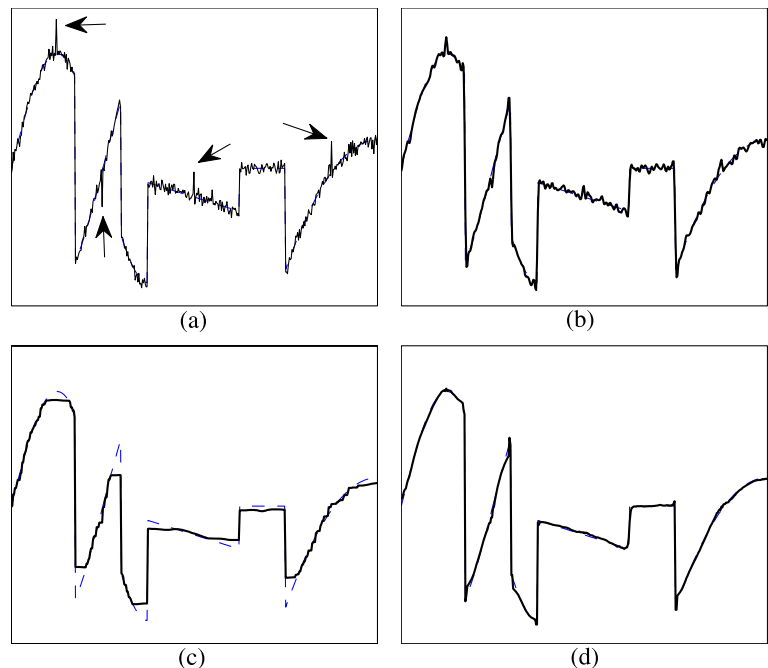
- The denoising property becomes better.
- The approximation becomes less affected by outliers.
- The local noisy artifact is eliminated.

These properties are due to the fact that the minimization of the TV norm eliminates small scaled structures very fast from the image.

As the outliers are eliminated very fast by the TV minimization, the local noisy artifact is greatly reduced and the approximation to the noise-free image gets better.

Figure 2 illustrates the fact that the incorporation of the total variation minimizing term reduces the influence of outliers. Figure 2(a) shows a noisy 1-D profile. The noisy profile is generated by adding Gaussian type noise to all positions, and some outlier noise to some positions (indicated

Fig. 2 Comparison of denoising properties on a 1-D profile (a) Noisy (b) MLS denoised (c) TV denoised (d) MLS-TV denoised



by arrows in the figure) of the original profile. Figure 2(b) is the result of applying the conventional MLS to Fig. 2(a) which shows that the large noise results in local noisy artifacts. Figure 2(c) shows the denoised result using the TV model which reveals also the staircasing artifact. Figure 2(d) shows the result of applying the MLS-TV to the noisy 1-D profile. Compared with Fig. 2(b), the local noisy artifact is eliminated with the proposed scheme, while effectively removing the noise.

Experimental results on images verifying the above properties will be given in the experimental section.

3.2 Contribution of the MLS-TV Model to the TV Model

As explained in Sect. 2, the minimization of the total variation norm tries to approximate the image to an approximation of first order. The fidelity term prevents the image from becoming totally flat, but cannot preserve the order of approximation, since there exist no inherent constraints in the u function that prevents the order from decreasing.

In contrast, the p function in the proposed model is constrained to have a pre-defined order due to the constraint to lie in the polynomial space Π_m . Therefore, the proposed model has the following advantageous properties over the conventional TV model.

- An approximation of any pre-defined order can be obtained.
- The scheme is less sensitive to the parameter deciding the strength of total variation minimization.

Here, one important fact is that a higher order approximation is obtained without varying the total variation minimization term, i.e., we can control the order of the solution without changing the regularization term. For example, if we want to achieve an approximation of degree 2 (which is equal to an approximation of order 3), we just let the solution p in (12) to lie in the polynomial space Π_2 . The constraint to lie in a certain polynomial space makes the solution also less sensitive to the parameter μ in (12) which decides the strength of competition between the regularization term and fidelity term in (12). This is due to the fact that the number of basis used in the reconstruction of p has a larger influence than the parameter μ . Furthermore, since the proposed scheme still uses the total variation minimization, it keeps the property of effective removing of outliers.

3.3 Iterative Algorithm to Solve the MLS-TV Model

In this section, we explain the numerics to obtain the minimizer $p(\mathbf{r})$ in (12). Since we construct p in Π_L of dimension m , we can write p as a linear combination of the basis $\{p_\ell, \ell = 1, \dots, m\}$:

$$p(\mathbf{x}) := \sum_{\ell=1}^m c_\ell p_\ell(\mathbf{x}). \tag{13}$$

Therefore, to find the minimizer $p(\mathbf{r})$ we have to obtain the following coefficients of the polynomial p ,

$$\mathbf{c} = \{c_\ell : \ell = 1, \dots, m\}.$$

A first try to obtain the coefficients might be to employ the energy function with variable \mathbf{c}

$$\mathcal{E}(\mathbf{c}) := \frac{\mu}{2} \|\mathbf{p} - \mathbf{f}\|_{\theta, \mathbf{x}}^2 + \|\nabla \mathbf{p}\|_1 \tag{14}$$

and try to find the coefficients by letting the derivative of the function $\mathcal{E}(\mathbf{c})$ equal to zero:

$$\frac{\partial \mathcal{E}(\mathbf{c})}{\partial c_\ell} = 0, \quad \ell = 1, \dots, m. \tag{15}$$

Here, $\mathbf{p} := \{p(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$ and $\mathbf{f} := \{f(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$ denote the evaluated values of the polynomial $p(\mathbf{x})$ at \mathbf{x}_i and the vectors of the data points in $\mathcal{N}(\mathbf{x})$, respectively, and thus \mathbf{p} depends on \mathbf{c} .

However, due to the L_1 norm, a linear system for the coefficients $\mathbf{c} = \{c_\ell : \ell = 1, \dots, m\}$ cannot be obtained. Therefore, to obtain a linear system for the coefficients, we exploit splitting methods. It will be shown later that using splitting methods we not only get a linear system but also an explicit form for \mathbf{c} . This is in contrast with the total variation minimization problem, where a linear system has to be solved, for example, by Gauss-Seidel iteration. We might employ any splitting method that results in a linear system for \mathbf{c} , but here we use the split Bregman iteration algorithm [30], which was introduced to solve the L_1 regularized optimization problem in Sect. 2.2, due to its simplicity.

By placing the iterative minimization scheme into our problem (12), we get the following iteration steps for each \mathbf{x} :

$$\begin{aligned} (1) \quad p^{k+1}(\mathbf{x}) &= \min_p \left\{ \sum_{n=1}^N \frac{\mu}{2} \|p(\mathbf{x}_n) - f(\mathbf{x}_n)\|_2^2 \theta(\mathbf{x}, \mathbf{x}_n) \right. \\ &\quad \left. + \frac{\lambda}{2} \|d^k(\mathbf{x}_n) - \nabla p(\mathbf{x}_n) - b^k(\mathbf{x}_n)\|_2^2 \theta(\mathbf{x}, \mathbf{x}_n) \right\} \\ (2) \quad d^{k+1}(\mathbf{x}) &= \min_d \|d\|_1 + \frac{\lambda}{2} \|d - \nabla p^{k+1}(\mathbf{x}) - b^k(\mathbf{x})\|_2^2 \\ &= \text{shrink}(\nabla p^{k+1}(\mathbf{x}) + b^k(\mathbf{x}), 1/\lambda) \\ (3) \quad b^{k+1}(\mathbf{x}) &= b^k(\mathbf{x}) + \nabla p^{k+1}(\mathbf{x}) - d^{k+1}(\mathbf{x}), \end{aligned} \tag{16}$$

where $\mathbf{x}_n \in \mathcal{N}(\mathbf{x})$, and k and $k + 1$ indicate the current and the next iteration level. The following notations are helpful for further discussion of the suggested method. From now on, any bolded character means a vector or matrix. Moreover, for a given function g and a weight function θ , the

weighted norm for a vector $\mathbf{g} := \{g(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$ is defined by

$$\|\mathbf{g}\|_{\theta, \mathbf{x}}^2 = \sum_{n=1}^N \|g(\mathbf{x}_n)\|_2^2 \theta(\mathbf{x}, \mathbf{x}_n).$$

We are now ready to explain the specific algorithm of the proposed scheme.

• **Step 1:** Let

$$\mathbf{f} := \{f(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$$

be the vector of the image brightness values at the data points \mathbf{x}_i in $\mathcal{N}(\mathbf{x})$ and

$$\mathbf{p} := \{p(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$$

be the vector of the evaluated values of the polynomial p at the data points \mathbf{x}_n in $\mathcal{N}(\mathbf{x})$. In other words, \mathbf{p} is the vector form of the polynomial p . Furthermore, let $\partial_x \mathbf{p}$ and $\partial_y \mathbf{p}$ be the vectors of the partial derivatives of $p(\mathbf{x}_i)$ with respect to x and y , respectively:

$$\partial_x \mathbf{p} := \{\partial_x p(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$$

$$\partial_y \mathbf{p} := \{\partial_y p(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\},$$

and let

$$\mathbf{b}_x := \{b_x^k(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$$

$$\mathbf{b}_y := \{b_y^k(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$$

$$\mathbf{d}_x := \{d_x^k(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\}$$

$$\mathbf{d}_y := \{d_y^k(\mathbf{x}_n) : \mathbf{x}_n \in \mathcal{N}(\mathbf{x})\},$$

where $b_x^k(\mathbf{x}_n)$ and $b_y^k(\mathbf{x}_n)$ are the x and y components of $b^k(\mathbf{x}_n)$, respectively, and likewise for $d_x^k(\mathbf{x}_n)$ and $d_y^k(\mathbf{x}_n)$. Then the first step (1) of the proposed model can be written in vector form as

$$\begin{aligned} \operatorname{argmin}_{p \in \Pi_m} & \frac{\mu}{2} \|\mathbf{p} - \mathbf{f}\|_{\theta, \mathbf{x}}^2 + \frac{\lambda}{2} \|\mathbf{d}_x - \partial_x \mathbf{p} - \mathbf{b}_x\|_{\theta, \mathbf{x}}^2 \\ & + \frac{\lambda}{2} \|\mathbf{d}_y - \partial_y \mathbf{p} - \mathbf{b}_y\|_{\theta, \mathbf{x}}^2. \end{aligned} \tag{17}$$

Since we construct p in Π_L of dimension m , we can write p as a linear combination of the basis $\{p_\ell, \ell = 1, \dots, m\}$:

$$p(\mathbf{x}) := \sum_{\ell=1}^m c_\ell p_\ell(\mathbf{x}). \tag{18}$$

Therefore, we see that (17) is a least squares problem that can be solved for the following coefficients of the polynomial p ,

$$\mathbf{c} = \{c_\ell : \ell = 1, \dots, m\}.$$

To this end, let us employ the energy function with variable \mathbf{c} ,

$$\begin{aligned} \mathcal{E}(\mathbf{c}) := & \frac{\mu}{2} \|\mathbf{p} - \mathbf{f}\|_{\theta, \mathbf{x}}^2 + \frac{\lambda}{2} \|\mathbf{d}_x - \partial_x \mathbf{p} - \mathbf{b}_x\|_{\theta, \mathbf{x}}^2 \\ & + \frac{\lambda}{2} \|\mathbf{d}_y - \partial_y \mathbf{p} - \mathbf{b}_y\|_{\theta, \mathbf{x}}^2, \end{aligned} \tag{19}$$

which is indeed a quadratic polynomial of the variables c_ℓ with $\ell = 1, \dots, m$. Then, the coefficient vector \mathbf{c} satisfying (19) can be found by calculating the derivative of the function $\mathcal{E}(\mathbf{c})$ and setting this equal to zero:

$$\frac{\partial \mathcal{E}(\mathbf{c})}{\partial c_\ell} = 0, \quad \ell = 1, \dots, m. \tag{20}$$

In order to find the solution of this linear system, we need to express (20) in a matrix form. For this, we denote by \mathbf{E} , \mathbf{E}_x and \mathbf{E}_y the matrices which (n, ℓ) components are defined as:

$$\mathbf{E}(n, \ell) := p_\ell(\mathbf{x}_n), \quad \mathbf{x}_n \in \mathcal{N}(\mathbf{x}), \quad \ell = 1, \dots, m$$

$$\mathbf{E}_x(n, \ell) := \partial_x p_\ell(\mathbf{x}_n), \quad \mathbf{x}_n \in \mathcal{N}(\mathbf{x}), \quad \ell = 1, \dots, m \tag{21}$$

$$\mathbf{E}_y(n, \ell) := \partial_y p_\ell(\mathbf{x}_n), \quad \mathbf{x}_n \in \mathcal{N}(\mathbf{x}), \quad \ell = 1, \dots, m,$$

where ∂_x and ∂_y are the partial derivative operators, i.e., $\partial_x = \frac{\partial}{\partial x}$ and $\partial_y = \frac{\partial}{\partial y}$. Moreover, recalling that the given weight function θ is non-zero for any $(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^2 \times \mathbb{R}^2$, let \mathbf{D}_θ indicate the $N \times N$ diagonal matrix whose (n, n) -component is defined by

$$\mathbf{D}_\theta(n, n) := \theta(\mathbf{x}_n, \mathbf{x}), \quad \mathbf{x}_n \in \mathcal{N}. \tag{22}$$

Then, using these notations, a simple calculation reveals that the condition (20) is equivalent to the following linear system:

$$\begin{aligned} \mathbf{0} = & \mu(\mathbf{c}\mathbf{E}^T - \mathbf{f})\mathbf{D}_\theta\mathbf{E} \\ & + \lambda(\mathbf{c}\mathbf{E}_x^T - (\mathbf{d}_x - \mathbf{b}_x))\mathbf{D}_\theta\mathbf{E}_x \\ & + \lambda(\mathbf{c}\mathbf{E}_y^T - (\mathbf{d}_y - \mathbf{b}_y))\mathbf{D}_\theta\mathbf{E}_y. \end{aligned}$$

This leads to the identity

$$\begin{aligned} \mathbf{c} = & (\mu\mathbf{f}\mathbf{D}_\theta\mathbf{E} + \lambda(\mathbf{d}_x - \mathbf{b}_x)\mathbf{D}_\theta\mathbf{E}_x + \lambda(\mathbf{d}_y - \mathbf{b}_y)\mathbf{D}_\theta\mathbf{E}_y) \\ & \cdot (\mu\mathbf{E}^T\mathbf{D}_\theta\mathbf{E} + \lambda\mathbf{E}_x^T\mathbf{D}_\theta\mathbf{E}_x + \lambda\mathbf{E}_y^T\mathbf{D}_\theta\mathbf{E}_y)^{-1}. \end{aligned} \tag{23}$$

We see that the solution \mathbf{c} is expressed explicitly, and therefore there is no need to solve a linear equation, which is in contrast with the TV minimization case. Each column of the matrices \mathbf{E} , \mathbf{E}_x and \mathbf{E}_y are composed of the polynomial basis and their partial derivatives. Therefore, it is easy to see that the matrix $[\mathbf{E}; \mathbf{E}_x; \mathbf{E}_y]$ is a full rank matrix. It is shown in [35], that for the conventional MLS problem we get the system matrix $\mu\mathbf{E}^T\mathbf{D}_\theta\mathbf{E}$, and that this matrix is invertible if \mathbf{E} is a full rank matrix and \mathbf{D}_θ is diagonal. Although it is not straightforward, following the

same procedure, it can be shown that the system matrix $\mu \mathbf{E}^T \mathbf{D}_\theta \mathbf{E} + \lambda \mathbf{E}_x^T \mathbf{D}_\theta \mathbf{E}_x + \lambda \mathbf{E}_y^T \mathbf{D}_\theta \mathbf{E}_y$ becomes invertible if $[\mathbf{E}; \mathbf{E}_x; \mathbf{E}_y]$ is a full rank matrix. Therefore, the inverse matrix always exists.

We see that (17) can be solved if the vectors \mathbf{d} and \mathbf{b} , and the initial noisy image \mathbf{f} are given. We initialized the vectors as $\mathbf{d} = \mathbf{0}$ and $\mathbf{b} = \mathbf{0}$. Finally, after obtaining this coefficient vector \mathbf{c} , our solution p^{k+1} is obtained by

$$p^{k+1} := \sum_{\ell=1}^m c_\ell p_\ell.$$

• **Step 2:** After p^{k+1} has been obtained in step 1, we can update d^{k+1} . To update $d^{k+1} := (d_x^{k+1}, d_y^{k+1})$ in step 2, we explicitly solve the minimization problem using the shrinkage operator used in [57] which is defined in Sect. 2.2, and simply compute

$$d^{k+1} = \text{shrink}(\nabla p^{k+1} + b^k, 1/\lambda) = \begin{cases} \nabla p^{k+1} + b^k - 1/\lambda \text{sgn}(\nabla p^{k+1} + b^k) & \text{if } |\nabla p^{k+1} + b^k| \geq 1/\lambda, \\ 0 & \text{if } |\nabla p^{k+1} + b^k| < 1/\lambda, \end{cases} \quad (24)$$

where $\text{sgn}(\cdot)$ is the sign operator.

• **Step 3:** After ∇p^{k+1} and d^{k+1} have been computed, the update of b^{k+1} is done using the same update method as in the Bregman algorithm.

$$b^{k+1} = b^k + \nabla p^{k+1} - d^{k+1}. \quad (25)$$

All the three steps have to be iteratively performed for each pixel $\mathbf{x} \in \Omega$, which gets us all the local polynomial functions $p(\mathbf{x}) = p_{\mathbf{x}}(\mathbf{x})$ for all $\mathbf{x} \in \Omega$. Then, finally, the overall MLS approximation function Lf is obtained by

$$Lf(\mathbf{x}) = p(\mathbf{x}) = p_{\mathbf{x}}(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega. \quad (26)$$

3.4 Other Combinations Using the Proposed Computational Framework

Other combinations of the least squares methods and the regularized variational models can be implemented using the computational framework explained in Sect. 3.3. For example, the least squares method with directional kernel [52] and the g-norm regularizer [37] can be put together with the proposed computational framework in the same manner as the MLS-TV model. Furthermore, we can also use other basis than polynomial basis to reconstruct p , for example, exponential basis [47, 54]. This will result in denoising models with other properties.

4 Experimental Results

We compared the denoising results of the proposed scheme with the conventional TV based denoising [45], the conven-

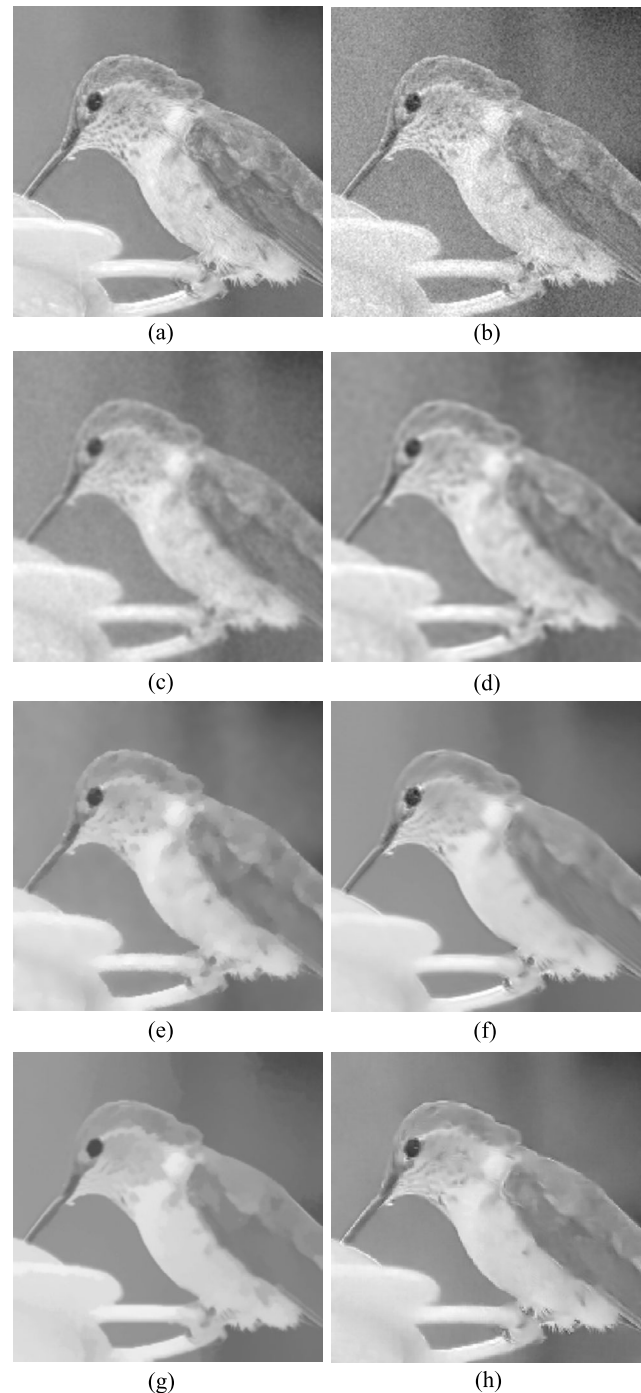


Fig. 3 Comparing the denoised results with the ‘Bird’ image (a) original (b) noisy ($\sigma = 12.75$, PSNR = 26.0636 dB) (c) denoised with MLS (PSNR = 28.2529 dB) (d) denoised with K-R (PSNR = 28.3158 dB) (e) denoised with TV (PSNR = 29.5978 dB) (f) denoised with NLM (PSNR = 29.7249 dB) (g) denoised with NLTV (PSNR = 28.0798 dB) (h) denoised with MLS-TV (PSNR = 30.5168 dB)

Table 1 PSNR comparison of the denoised images by 8 different denoising algorithms, where the noise is derived from a Gaussian distribution with standard deviation $\sigma = 15.3$ (and $\sigma = 20.4$ for the ‘Cameraman’ image with PSNR = 22.2497)

	Barbara (24.4297)	Cameraman (24.4535)	Cameraman (22.2497)	House (24.4291)	Lena (24.4245)	Pirate (24.4642)
K-R [52]	27.2142	32.2122	24.1324	35.2231	31.2146	28.7599
MLS [35]	29.8410	32.6258	28.9740	34.3328	32.1342	30.7242
NLM [11]	30.7963	32.8516	27.8477	35.7475	32.2955	29.6953
NL-SS [28]	30.4070	32.5087	28.8289	35.0515	32.2902	30.0761
NLM-SAP [22]	31.3233	31.9153	29.4391	34.0924	32.2028	30.0910
TV ($\mu = 5$) [45]	23.2493	27.8844	27.0242	31.1925	27.7914	26.0073
TV ($\mu = 10$)	24.6651	30.6648	27.1164	33.5688	30.3118	28.2854
TV ($\mu = 15$)	26.2151	32.2118	27.2562	34.7114	31.6496	29.6941
TV ($\mu = 20$)	27.4128	32.8499	27.5312	34.8779	32.1867	30.7167
NLTV ($\mu = 5$) [61]	26.4092	29.4682	27.9632	32.6588	28.8038	26.7343
NLTV ($\mu = 10$)	27.9442	30.6691	28.1291	33.5527	30.0034	27.8442
NLTV ($\mu = 15$)	29.0939	31.6607	28.5713	34.2373	30.9416	28.7855
NLTV ($\mu = 20$)	29.9209	32.4654	28.6154	34.7770	31.6599	29.5691
MLS-TV ($\mu = 5$)	31.8894	32.0233	27.7223	33.9455	31.8842	29.9223
MLS-TV ($\mu = 10$)	32.1009	32.9362	28.3934	34.9046	32.1009	30.0210
MLS-TV ($\mu = 15$)	32.5240	33.3529	28.8635	35.4058	32.6126	30.4844
MLS-TV ($\mu = 20$)	32.4647	33.5713	29.2147	35.4231	32.6219	30.5332

Fig. 4 Showing some details of Fig. 3 for the denoised results using the TV and the MLS-TV models (a) TV (b) MLS-TV

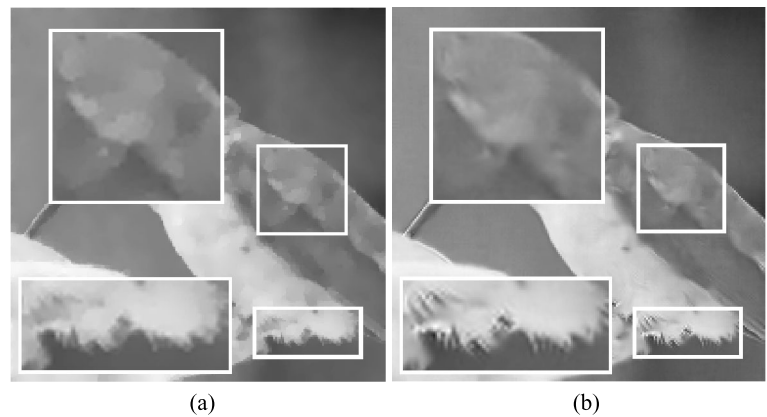
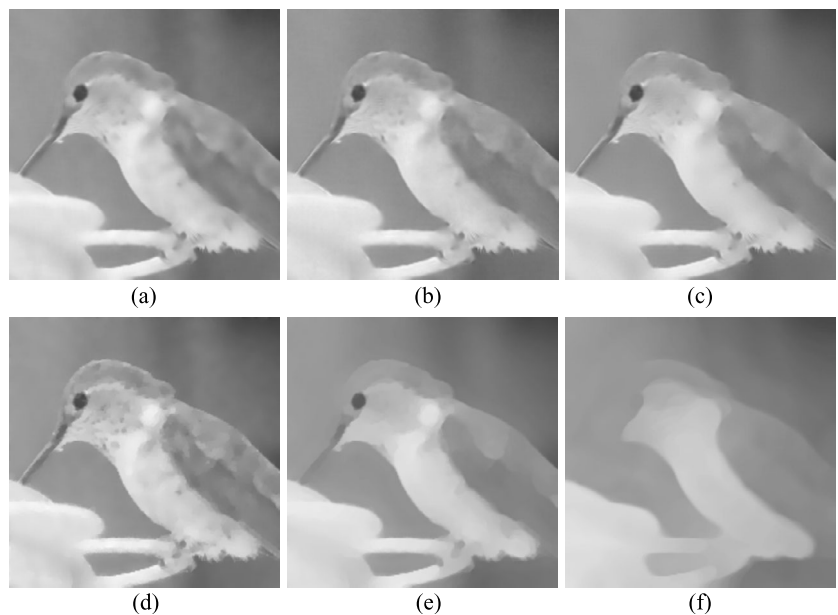


Fig. 5 Upper row: MLS-TV (a) $\mu = 15$ (b) $\mu = 5$ (c) $\mu = 1$, Lower row: TV (d) $\mu = 15$ (e) $\mu = 5$ (f) $\mu = 1$



tional MLS based denoising [35], the *nonlocal TV* (NLTV) based denoising [61], the *kernel regression* (K-R) based denoising [52], *nonlocal mean* (NLM) based denoising [11], the *nonlocal scale-shape* (NL-SS) [28], and the *nonlocal method with shape-adaptive patches* (NLM-SAP) [22]



Fig. 6 Comparing the denoised results with the ‘Barbara’ image (a) original (b) noisy ($\sigma = 15.3$, PSNR = 24.4437 dB) (c) denoised with MLS (PSNR = 29.8410 dB) (d) denoised with K-R (PSNR = 27.2142 dB) (e) denoised with TV (PSNR = 26.2151 dB) (f) denoised with NLM (PSNR = 30.7963 dB) (g) denoised with NLTV (PSNR = 29.0939 dB) (h) denoised with MLS-TV (PSNR = 32.5240 dB)

methods. The TV and the NLTV methods are implemented using the split Bregman method. The NL-SS uses the nonlocal weighted linear diffusion process, while the NLM-SAP

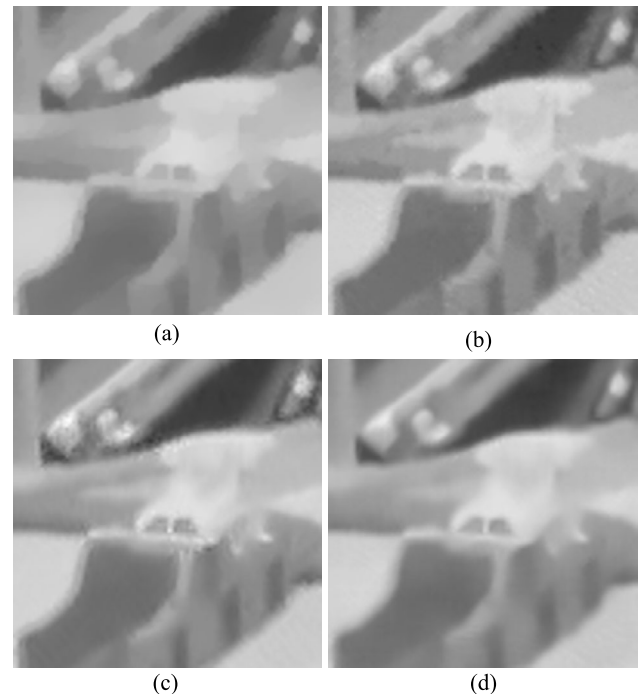


Fig. 7 Enlarged versions of the denoised results in Fig. 6. (a) TV (b) NLTV (c) NLM (d) MLS-TV

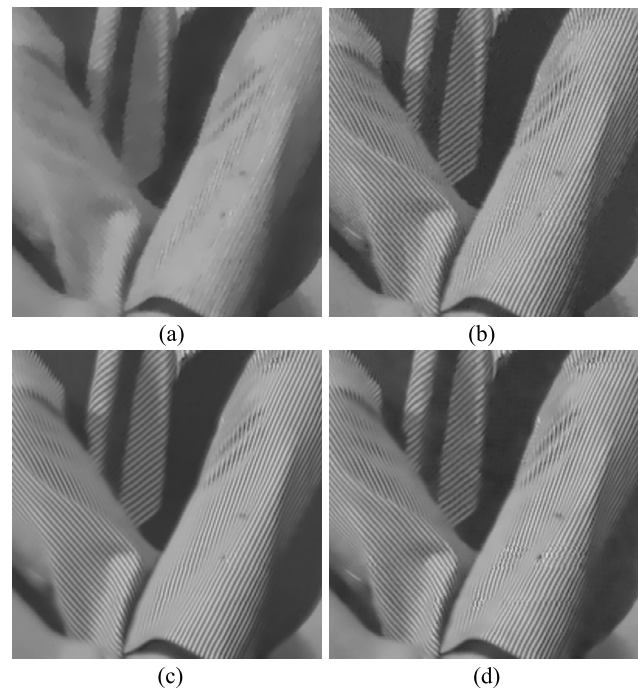


Fig. 8 Enlarged versions of the denoised results in Fig. 6. (a) TV (b) NLTV (c) NLM (d) MLS-TV

uses more general shapes for the patch in NLM instead of the usual square patches.

The easiness in choosing the parameters is also an important factor to measure the goodness of the algorithm. While the proposed scheme is not so sensitive to the parameters, the other schemes are sensitive to the choice of parameters, especially the K-R based method. This is shown in Fig. 5 and also in Table 1, where it can be observed that the PSNR values vary much for different parameters with other algorithms, while for the proposed one the PSNR values are similar.

Figure 3 compares the denoising results on the ‘Bird’ image with Gaussian noise, where the PSNR is 26.0636 dB, and the noise is generated from a zero mean Gaussian distribution with $\sigma = 12.75$, where σ is the standard deviation. The local density and the global smoothing parameters [52] used in the K-R method are 0.1 and 0.5, respectively, and the structural sensitivity is 0.1. The width of Gaussian kernel in the NLM is 0.06. The fidelity parameters (μ) for the TV, NLTV, MLS-TV are all set to 15 (with $\lambda = 5$ in (16)), and the number of data used in the reconstruction of the local polynomial approximation (N in (16)) is 49 for both the MLS and the MLS-TV models. Here, the parameter λ is the Bregman parameter which appears in the Bregman formu-

lation of the model, and has an effect on the convergence speed of the algorithm. The degree of polynomials and the width of Gaussian are 3 and 0.3, respectively, for both the MLS and the MLS-TV.

As can be seen in the backgrounds of Figs. 3(c) and 3(d), the conventional MLS and the K-R based denoising show local noisy artifacts due to the strong noise which affects the neighborhood of kernel size. In contrast, with the proposed scheme, the noisy artifact is effectively removed due to the TV minimizing term. Compared with the conventional TV based denoising scheme, the proposed scheme produces a higher order approximation to the intensity function, and the staircasing artifact does not appear in smooth regions. The visual quality of the denoised image seems to be better even than those removed by the conventional NLTV and the NLM based denoising schemes.

Figure 4 shows some details of Figs. 3(e) and 3(h) to verify that the approximation using the proposed scheme is of higher order than that of the TV model.

Figure 5 shows that the proposed scheme is not so sensitive to the parameter deciding the strength of fidelity (μ in (12)) as the TV model is. This is due to the additional constraint in the reconstruction of the p function. The insensitivity to the parameter makes the scheme stable.

Fig. 9 Comparing the denoised results with the ‘Cameraman’ image (a) noisy ($\sigma = 20.4$, PSNR = 22.2497 dB) (b) denoised with MLS (PSNR = 28.9740 dB) (c) denoised with K-R (PSNR = 24.1324 dB) (d) denoised with TV (PSNR = 27.2562 dB) (e) denoised with NLM (PSNR = 27.8477 dB) (f) denoised with NLTV (PSNR = 28.5713 dB) (g) denoised with NLM-SAP (PSNR = 29.4391 dB) (h) denoised with NL-SS (PSNR = 28.8289 dB) (i) denoised with MLS-TV (PSNR = 28.8635 dB)

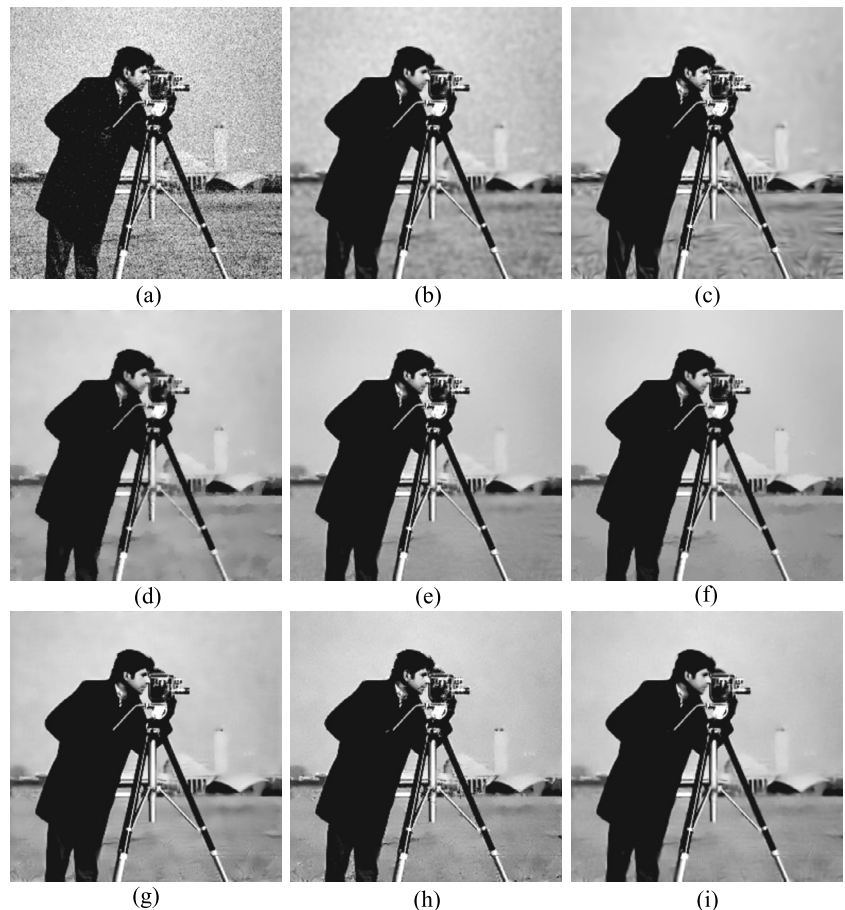


Fig. 10 Enlarged versions of the denoised results in Fig. 9 (a) noisy ($\sigma = 20.4$, PSNR = 22.2497 dB) (b) denoised with MLS (PSNR = 28.9740 dB) (c) denoised with K-R (PSNR = 24.1324 dB) (d) denoised with TV (PSNR = 27.2562 dB) (e) denoised with NLM (PSNR = 27.8477 dB) (f) denoised with NLTV (PSNR = 28.5713 dB) (g) denoised with NLM-SAP (PSNR = 29.4391 dB) (h) denoised with NL-SS (PSNR = 28.8289 dB) (i) denoised with MLS-TV (PSNR = 28.8635 dB)

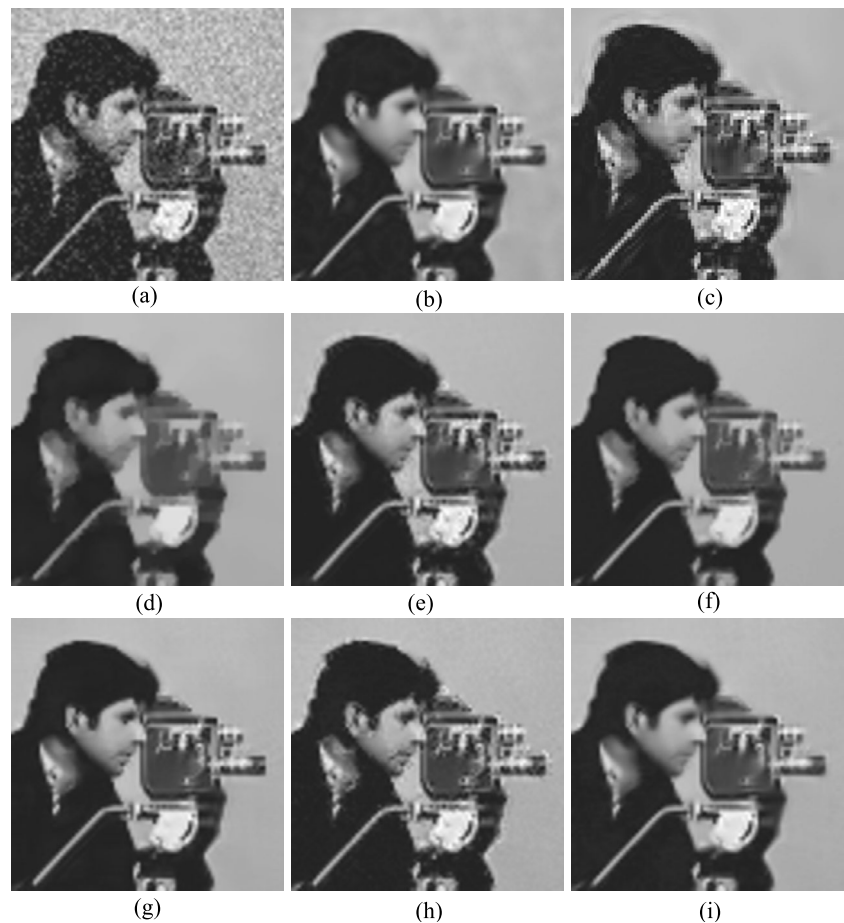


Figure 6 compares the denoising results with the ‘Barbara’ image, which contains image structures with large oscillation. We used the following parameters in the denoising of the ‘Barbara’ image. The used in the K-R method are same as used in the ‘Bird’ image. The width of the Gaussian kernel in the NLM is 0.08. The fidelity parameters (μ) for the TV, NLTV, MLS-TV are 10, 20, and 10, respectively, with $\lambda = 5$. The number of data (N), the degree of polynomials, and the width of Gaussian are 49, 3, 0.3, respectively, for both the MLS and the MLS-TV. Again, the staircase artifact and the *noise halo* artifact can be observed with the results of the TV and the NLM schemes, respectively, especially in the enlarged figures in Fig. 7. With the proposed scheme, these artifacts are suppressed. The proposed scheme can also preserve the stripes in the ‘Barbara’ image, as the NLM scheme does, as can be seen in Fig. 8.

Figure 9 compares the denoising results with the ‘Cameraman’ image, which has a stronger Gaussian noise ($\sigma = 20.4$, PSNR = 22.2497 dB) than the ‘Bird’ and ‘Barbara’ image. For this, we compare the result with two more non-local approach, NLM-SAP and NL-SS. The width of Gaussian in the both methods is 0.08. We used 15 different shapes for the patch in NLM-SAP and $\Delta t = 1/14$ for the steep-

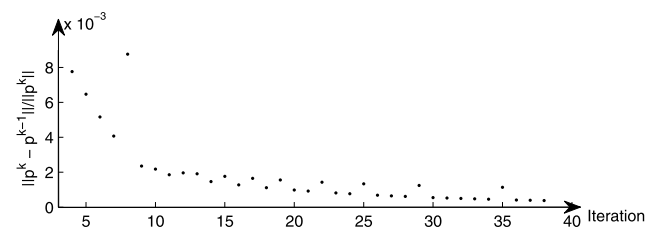


Fig. 11 Change of the $\|Lf^{k+1} - Lf^k\| / \|Lf^k\|$ value according to the number of iterations

est descent in NL-SS. The other parameters are the same as with the ‘Barbara’ image for the other methods. Due to the strong noise, the results with the conventional MLS and the K-R based denoising suffer from significant local noisy artifacts. In removing the strong noise, the staircase artifact of the TV and the NLTV scheme and the *noise halo* artifact of the NLM scheme and NL-SS scheme, a typical artifact of the NLM scheme which is due to an abrupt lack of redundancy around edges, become apparent. These artifacts can be better observed in the enlarged images in Fig. 10. The NLM-SAP has the largest PSNR value, but artifacts can be seen in the background. In contrast, the MLS-TV scheme has a better visual quality as can be seen in Figs. 9(i) and 10(i).

Table 1 further compares the PSNR values of the denoising results on five different 512×512 noisy images, where the noise was generated with a zero mean Gaussian distribution with $\sigma = 15.3$. For the ‘Cameraman’ image, we also tabled the results from Fig. 9, which has a stronger noise ($\sigma = 20.4$, PSNR = 22.2497 dB). Especially, the NLTV, TV, and MLS-TV methods are implemented with four dif-

ferent fidelity parameters for each image. The width of the Gaussian kernels used in the NLM, NLM-SAP and NL-SS methods is 0.06, and the other parameters are the same as those used in Fig. 10. We see that the proposed MLS-TV method has a good denoising result compared to the other algorithms, when the noise is not too strong.

We used a stopping criterion to terminate the iteration of the proposed scheme. The iteration stops if $\|Lf^{k+1} - Lf^k\|/\|Lf^k\| < 10^{-3}$, where Lf^k denotes the approximated image at the current step, and Lf^{k+1} denotes that at the next step. To show the convergence property of the proposed scheme, we draw a convergence plot of the proposed scheme in Fig. 11. Figure 11 shows that the value $\|Lf^{k+1} - Lf^k\|/\|Lf^k\| < 10^{-3}$ converges.

Figure 12 shows the plot of the PSNR value and the total iteration number versus the fidelity parameter value, where the PSNR values corresponds to the resulted images obtained with the stopping criterion $\|Lf^{k+1} - Lf^k\|/\|Lf^k\| <$

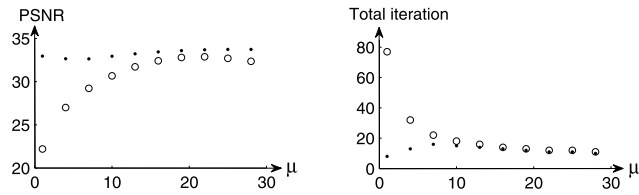


Fig. 12 Change of the PSNR and the total iteration according to the fidelity parameter (μ). The small dots corresponds to the proposed method and the large dot to the TV method

Fig. 13 Showing the denoising results on Fig. 9(a) with the proposed method using a degree 0 (first column), degree 3 (second column), degree 4 (third column) approximation, and a neighborhood of size 5×5 (first row), 7×7 (second row), 9×9 (third row), 11×11 (fourth row). The parameters $\mu = 15$, $\lambda = 5$ have been used for all images



10^{-3} for the TV method and the proposed method. This plot shows the fact that the proposed method is less sensitive to the fidelity parameter than the TV method.

Figure 13 shows how the denoised result of the proposed method depends on the degree of the approximation and on $\mathcal{N}(\mathbf{x})$, the size of the neighborhood. It can be seen that the denoised results are similar for the images of degree 3 and degree 4 approximations, since the difference in the approximation error between the degree 3 and the degree 4 approximations is small.

All the algorithms are implemented with the Matlab program (version R2011a, 64-bit) using only Matlab commands on an Intel Core2 CPU (2.66 GHz) and 8 GB memory. With this configuration, the running time was about 180 seconds on a 256×256 image with the proposed scheme. The running time is large compared with the TV method which took about 0.8 seconds. The K-R, MLS, NLTV, NLM, NLM-SAP, and the NL-SS methods took 47, 23, 1, 60, 36, 518 seconds, respectively. However, the code corresponding to the proposed scheme was not optimized, and there is the possibility that the running time of the proposed scheme can be significantly reduced after optimization.

5 Conclusion

In this paper, we have proposed a computational framework which can put regularity based variational methods and least squares methods together. As an example, we showed how the total variation minimization can be incorporated into the moving least squares method, which resulted in a denoising scheme that can overcome the drawbacks of both the conventional total variation based and the moving least squares based denoising schemes. The proposed framework can be used to further incorporate several other regularization terms which have been proposed in the variational approaches into the least squares methods. Meanwhile, methods adopted from the least squares framework such as the usage of steering kernels can also be incorporated in the proposed framework. The proposed framework has opened possibilities for numerous schemes which can be derived in the same manner as the proposed MLS-TV scheme.

Acknowledgements This work was supported by the Basic Science Research Programs 2012R1A1A2004518 (J. Yoon), 2010-0011689 (Y. Lee), 2010-0006567 (S. Lee), and the Priority Research Centers Program 2009-0093827 (Y. Lee and J. Yoon) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology.

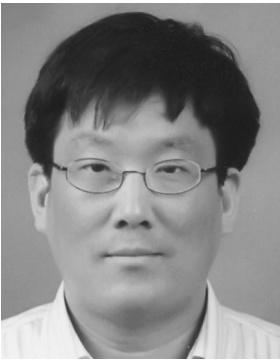
References

- Aubert, G., Kornprobst, P.: *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Applied Mathematical Sciences, vol. 147. Springer, Berlin (2002)
- Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
- Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **18**(11), 2419–2434 (2009)
- Blomgren, P., Chan, T., Mulet, P., Wong, C.K.: Total variation image restoration: numerical methods and extensions. In: *IEEE ICIP*, pp. 384–387 (1997)
- Bose, N.K., Ahuja, N.A.: Superresolution and noise filtering using moving least squares. *IEEE Trans. Image Process.* **15**(8), 2239–2248 (2006)
- Bougleux, S., Elmoataz, A., Melkemi, M.: Local and nonlocal discrete regularization on weighted graphs for image and mesh processing. *Int. J. Comput. Vis.* **84**(2), 220–236 (2009)
- Breitkopf, P., Naceur, H., Passineux, A., Villon, P.: Moving least squares response surface approximation: formulation and metal forming applications. *Comput. Struct.* **83**(17–18), 1411–1428 (2005)
- Brox, T., Kleinschmidt, O., Cremers, D.: Efficient nonlocal means for denoising of textural patterns. *IEEE Trans. Image Process.* **17**(7), 1083–1092 (2008)
- Bruhn, A., Weickert, J.: Towards ultimate motion estimation: combining highest accuracy with real-time performance. In: *IEEE Int. Conference on Computer Vis.*, vol. 1, pp. 749–755 (2005)
- Bruhn, A., Weickert, J., Kohlberger, T., Schnörr, C.: A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *Int. J. Comput. Vis.* **70**(3), 257–277 (2006)
- Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **4**(2), 490–530 (2005)
- Caselles, V., Chambolle, A., Novaga, M.: The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model. Simul.* **6**(3), 879–894 (2007)
- Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**(1–2), 89–97 (2004)
- Chambolle, A., Lions, P.-L.: Image recovery via total variation minimization and related problems. *Numer. Math.* **76**(2), 167–188 (1997)
- Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011)
- Chan, T., Marquina, A., Mulet, P.: High-order total variation based image restoration. *SIAM J. Sci. Comput.* **22**(2), 503–516 (2000)
- Chan, T., Esedoglu, S., Park, F., Yip, A.: *Recent Developments in Total Variation Image Restoration*. Mathematical Models in Computer Vision. Springer, Berlin (2005)
- Chan, T., Esedoglu, S., Park, F.E.: Image decomposition combining staircase reduction and texture extraction. *J. Vis. Commun. Image Represent.* **18**(6), 464–486 (2007)
- Chan, T., Esedoglu, S., Park, F.: A fourth order dual method for staircase reduction in texture extraction and image restoration problems. In: *IEEE ICIP*, pp. 4137–4140 (2010)
- Chu, C.K., Glad, I.K., Godtliebsen, F., Marron, J.S.: Edge-preserving smoothers for image processing. *J. Am. Stat. Assoc.* **93**(442), 526–556 (1998)
- Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007)
- Deledalle, C.A., Duval, V., Salmon, J.: Non-local methods with shape-adaptive patches (NLM-SAP). *IMA J. Numer. Anal.* **43**(2), 103–120 (2012)
- Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program., Ser. A* **55**(3), 293–318 (1992)

24. Elad, M.: On the origin of the bilateral filter and ways to improve it. *IEEE Trans. Image Process.* **11**(10), 1141–1151 (2002)
25. Esser, E.: Applications of Lagrangian-based alternating direction methods and connections to split-Bregman. *UCLA CAM-Reports* 09-31 (2009)
26. Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.* **3**(4), 1015–1046 (2010)
27. Fenn, M., Steidl, G.: Robust local approximation of scattered data. In: *Geometric Properties from Incomplete Data*, vol. 31, pp. 317–334 (2006)
28. Gilboa, G., Osher, S.: Nonlocal linear image regularization and supervised segmentation. *Multiscale Model. Simul.* **6**(2), 595–630 (2007)
29. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. *Multiscale Model. Simul.* **7**(3), 1005–1028 (2008)
30. Goldstein, T., Osher, S.: The split Bregman method for L1 regularized problems. *SIAM J. Imaging Sci.* **2**(2), 323–343 (2009)
31. Gwosdek, P., Bruhn, A., Weickert, J.: Variational optic flow on the Sony PlayStation 3—accurate dense flow fields for real-time applications. *J. Real-Time Image Process.* **5**(3), 163–177 (2010)
32. Kamilov, U., Bostan, E., Unser, M.: Generalized total variation denoising via Augmented Lagrangian cycle spinning with Haar wavelets. In: *IEEE ICASSP*, pp. 909–912 (2012)
33. Kervrann, C., Boulanger, J.: Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *Int. J. Comput. Vis.* **79**(1), 45–69 (2008)
34. Lancaster, P., Salkauskas, K.: Surfaces generated by moving least squares methods. *Math. Comput.* **37**(155), 141–158 (1981)
35. Levin, D.: The approximation power of moving least-squares. *Math. Comput.* **67**(224), 1517–1531 (1998)
36. Lions, P.-L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16**(6), 964–979 (1979)
37. Meyer, Y.: Oscillating patterns in image processing and nonlinear evolution equations. In: *The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, vol. 22. AMS, Providence (2001)
38. Mirzaei, D., Schaback, R., Dehghan, M.: On generalized moving least squares and diffuse derivatives. *IMA J. Numer. Anal.* **32**(3), 983–1000 (2012)
39. Mrazek, P., Meickert, J., Bruhn, A.: On robust estimation and smoothing with spatial and tonal kernels. In: *Geometric Properties from Incomplete Data*, vol. 31, pp. 335–352 (2006)
40. Nayroles, B., Touzot, G., Villon, P.: Generalizing the finite element method: diffuse approximation and diffuse elements. *Comput. Mech.* **10**(5), 307–318 (1992)
41. Nikolova, M.: Local strong homogeneity of a regularized estimator. *SIAM J. Appl. Math.* **61**(2), 633–658 (2000)
42. Nikolova, M.: Weakly constrained minimization: application to the estimation of images and signals involving constant regions. *J. Math. Imaging Vis.* **21**(2), 155–175 (2004)
43. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.* **4**(2), 460–489 (2005)
44. Pizarro, L., Mrázek, P., Didas, S., Grewenig, S., Weickert, J.: Generalised nonlocal image smoothing. *Int. J. Comput. Vis.* **90**(1), 62–87 (2010)
45. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**(1–4), 259–268 (1992)
46. Sapiro, G.: *Geometric Partial Differential Equations and Image Processing*. Cambridge University Press, Cambridge (2001)
47. Schumaker, L.L.: *Spline Functions: Basic Theory*. Wiley, New York (1981)
48. Setzer, S.: Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In: *International Conf. Scale Space and Variational Methods in Computer Vis.*, vol. 5567, pp. 464–476. Springer, Berlin (2009)
49. Setzer, S.: Operator splitting, Bregman methods and frame shrinkage in image processing. *Int. J. Comput. Vis.* **92**(3), 265–280 (2011)
50. Setzer, S.: Infimal convolution regularizations with discrete L1-type functionals. *Commun. Math. Sci.* **9**(3), 797–827 (2011)
51. Stefan, W., Renaut, R.A., Gelb, A.: Improved total variation-type regularization using higher order edge detectors. *SIAM J. Imaging Sci.* **3**(2), 232–251 (2010)
52. Takeda, H., Farsiu, S., Milanfar, P.: Kernel regression for image processing and reconstruction. *IEEE Trans. Image Process.* **16**(2), 349–366 (2007)
53. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: *Proc. IEEE Int. Conf. Computer Vision*, pp. 839–846 (1998)
54. Unser, M., Blu, T.: Cardinal exponential splines: Part I: theory and filtering algorithms. *IEEE Trans. Signal Process.* **53**(4), 1425–1438 (2005)
55. Vogel, C.R., Oman, M.E.: Iterative methods for total variation denoising. *SIAM J. Sci. Comput.* **17**(1), 227–238 (1996)
56. Wang, C., Liu, Z.: Total variation for image restoration with smooth area protection. *J. Signal Process. Syst.* **61**(3), 271–277 (2010)
57. Wang, Y., Yin, W., Zhang, Y.: A fast algorithm for image deblurring with total variation regularization. *CAAM Technical Report* 07-10 (2007)
58. Winker, G., Aurich, V., Hahn, K., Martin, A., Rodenacker, K.: Noise reduction in images: some recent edge-preserving methods. *Pattern Recognit. Image Anal.* **9**(4), 749–766 (1999)
59. Wu, C., Tai, X.-C.: Augmented Lagrangian method, dual methods, and split Bregman iterations for ROF, vectorial TV and higher order models. *SIAM J. Imaging Sci.* **3**(3), 300–339 (2010)
60. Yoon, J., Lee, Y.: Nonlinear image upsampling method based on radial basis function interpolation. *IEEE Trans. Image Process.* **19**(10), 2682–2692 (2010)
61. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM J. Imaging Sci.* **3**(3), 253–276 (2010)
62. Zhang, X., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. *J. Sci. Comput.* **46**(1), 20–46 (2011)
63. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Reports* 08-34 (2008)



Yeon Ju Lee received the B.S., M.S., and Ph.D. degrees from the Dept. of Mathematics, Ewha W. University, Seoul, South Korea, in 2001, 2003, and 2007, respectively. She was a Postdoctoral Research Associate from 2007 to 2009 in the Department of Computer Sciences, University of Wisconsin-Madison. From 2009 to 2012, she was a Postdoctoral Scholar at the Department of Mathematical Sciences, KAIST, Daejeon, South Korea. She is now a Research Professor at the Institute of Mathematical Sciences, Ewha W. University. Her research interests are in the area of data approximation, signal/image processing, wavelet, and subdivision.



Sukho Lee received his BS, MS, and PhD in electronics engineering from Yonsei University, Korea, in 1993, 1998, and 2003, respectively. He worked as a researcher in the Impedance Imaging Research Center from 2003 to 2006 and was an assistant professor at Yonsei University from 2006 to 2008. He has been in the Department of Software Engineering at Dongseo University, Busan, Korea, since 2008, where he is now a professor. His current research interests include image and video filtering based

on PDEs, medical imaging, and computer vision.



Jungho Yoon received the Ph.D. degree in 1998 from the University of Wisconsin, Madison, USA. From 1998 to 2000, he held a position of Postdoctoral Research Associate at the Department of Mathematics of University of South Carolina, USA. From 2000 to 2001, he was a Postdoctoral Research Associate at the School of Mathematical Sciences of Tel-Aviv University, Israel. From 2001–2002, he was a Visiting Assistant Professor at the Department of Mathematics of Arizona

State University, USA. He is currently a Professor at the Department of Mathematics of Ewha Womans University, South Korea. His research interests include image processing, approximation, subdivision and wavelet.