Image Denoising Using Backward Stochastic Differential Equations

Dariusz Borkowski[1(✉)] and Katarzyna Jańczak-Borkowska[2]

[1] Faculty of Mathematics and Computer Science, Nicolaus Copernicus University,
Chopina 12/18, 87-100 Toruń, Poland
dbor@mat.umk.pl
[2] Institute of Mathematics and Physics, University of Science and Technology,
al. prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland
kaja@utp.edu.pl

Abstract. In this paper we explore the problem of reconstruction of RGB images with additive Gaussian noise. In order to solve this problem we use backward stochastic differential equations. The reconstructed image is characterized by smoothing noisy pixels and at the same time enhancing and sharpening edges. This novel look on the reconstruction is fruitful, gives encouraging results and can be successfully applied to denoising of high ISO images.

Keywords: Image denoising · Inverse problem · Backward stochastic differential equations

1   Introduction

The backward stochastic differential equations (BSDEs) were introduced by Pardoux and Peng [14] who proved the existence and uniqueness of their adapted solution under suitable assumptions. Independently, Duffie and Epstein [9,10] considered stochastic differential utilities in economics models, as solution of certain BSDEs. Since then, it has been widely recognized that BSDEs provide a useful framework for formulating many problems in mathematical finance [11]. These equations have also appeared to be useful for problems in stochastic control and differential games [13]. Many papers (for instance [15]) show the connections between BSDEs driven by a diffusion process and solutions of a large class of quasilinear parabolic and elliptic partial differential equations (PDEs). These results may be seen as a generalization of the celebrated Feynman-Kac formula. Through all these results, a formal dictionary between BSDEs and PDEs can be established, which suggests that existence and uniqueness results obtained on the one side should have their counterparts on the other side. In our case, we treat BSDEs as a starting point in the creation of reconstruction model of the noisy image. In our opinion a stochastic description is more intuitive than PDEs description. Moreover in order to solve the BSDE model we use tools of stochastic analysis giving us completely new methodology.

In image processing one can find some practical aspects [1–4] of BSDE-based applications. In [1] the problem of reconstruction of the noisy chromaticity is considered. Presented model of denoising is expressed in terms of Skorokhod problem associated with the solution of BSDE and an epsilon neighbourhood of two dimensional sphere. In that paper BSDE is driven by a trivial drift function ($f \equiv 0$). This means that presented equation is a model of forward filtering and has properties of smoothing filters. In [2] problems of reconstruction of the noisy grayscale image (smoothing filters) and enhancing of the blurred grayscale image (enhancing filters) are presented. Smoothing filter, similarly as in [1], models on anisotropic forward diffusion with BSDEs with $f \equiv 0$. Enhancing filters presented in [2] are driven by BSDEs with non trivial drift function and correspond to inverse heat equation. This equation is a model of backward filtering (not forward) and in consequence this model fails to edge enhancing of the noisy image. The paper [4] is a generalisation of [2] to images with values in $\mathbf{R}^n$. In papers [2,4] smoothing and enhancing filters were considered as two separable models while in [3] authors proposed the model of image reconstruction using backward stochastic diffusion associated with forward stochastic diffusion. This method combines forward anisotropic filtering in perpendicular to gradient direction and inverse anisotropic filtering in gradient direction, where effects of smoothing, enhancing and sharpening join. The article [3] focuses only on images with values in $\mathbf{R}^1$ (greyscale images).

In this paper we generalise the results from [3] to $\mathbf{R}^n$ valued images, in particular to colour images and apply the proposed method to denoising of high ISO images taken from digital cameras.

The paper is constructed as follows. Sections 2 and 3 contain basic ideas of filtering in terms of BSDEs taken from [3]. Section 4 provides new results to reconstruction of colour images. In Sect. 5 we give details of numerical approximation of the proposed method. Finally, in Sect. 6 experimental results to denoising of high ISO images taken from digital cameras are presented.

2   Continuous Model

A general model of the image reconstruction in terms of BSDE is the following [3]:

$$
\begin{cases}
X_t = x + \displaystyle\int_0^t \sigma(s, X_s)\, dW_s + K_t^{\overline{D}}, \quad t \in [0, T] \\[2mm]
Y_t = \xi + \displaystyle\int_t^T f(s, Y_s, X_s)\, ds - \int_t^T Z_s\, dW_s, \quad t \in [0, T]
\end{cases}
\tag{1}
$$

where ξ is some random variable which depends on the noisy image $u_0 : \overline{D} \to \mathbf{R}^n$ and $\{W_t\}_{t \in [0,T]}$ is two-dimensional Wiener process. $\{X_t\}_{t \in [0,T]}$ is a stochastic diffusion process with reflection with values in the domain of the image $\overline{D}$ and is driven by some function σ. The function σ is called the diffusion coefficient.

The process X characterizes the behaviour of the continuous time stochastic process X as an Itö integral. A heuristic interpretation is that in a small time interval of length σ the stochastic process X changes its value by an amount that is normally distributed with variance $\sigma(t, X_t)$ and is independent of the past behaviour of the process. This is so because the increments of a Wiener process are independent and normally distributed. The term $\{K_t^{\overline{D}}\}_{t \in [0,T]}$ is the minimal push needed to keep process X in $\overline{D}$. $\{Y_t\}_{t \in [0,T]}$ (the first component of the solution to the BSDE) is a backward stochastic diffusion process with values in the codomain of the image $\mathbf{R}^n$ and is driven by some function f. $\{Z_t\}_{t \in [0,T]}$, the second component of the solution to the BSDE determines the measurability of the process Y. The process Y is constructed so that it starts at $t = T$ until $t = 0$. A drift function f causes the correction of values of Y in expected strength and direction. The value of the process Y at time $t = 0$ is the reconstructed pixel $u(x)$. Note that Y_0 is a deterministic value.

In the case of image processing we have the following interpretation of this BSDE equation. For a fixed pixel x we consider a certain BSDE equation. The values of the process X determines pixels from domain of the image $\overline{D}$ which we will use in process reconstruction. We can say that this process determines neighbourhood of the pixel x (with irregular shape). The reconstructed value $u(x)$ is the sum of pixels from its neighbourhood multiplied by some weights. The weight values are determined by the process Y, a more specifically by the function f. Appropriate definition of the function f allows as to give weight values (also negative) which depend on direction and distance from reconstructed pixel.

3   Grayscale Images

In [3] the authors proposed the following method of the reconstruction of grayscale images:

$$
\begin{cases}
X_t = x + \int_0^t \sigma(s, X_s)\, dW_s + K_t^{\overline{D}}, & t \in [0,T], \\
Y_t = u_0(X_S) + \int_t^T c(s)(Y_s - u_0(X_s))ds - \int_t^T Z_s\, dW_s, & t \in [0,T],
\end{cases}
\tag{2}
$$

where $S < T$,

$$
\sigma(s, X_s) = \begin{bmatrix}
-\left(1 - \dfrac{c(s)}{c}\right) \dfrac{(G_\gamma * u_0)_{x_2}(X_s)}{|\nabla(G_\gamma * u_0)(X_s)|}, & \dfrac{c(s)}{c} \dfrac{(G_\gamma * u_0)_{x_1}(X_s)}{|\nabla(G_\gamma * u_0)(X_s)|} \\[2ex]
\left(1 - \dfrac{c(s)}{c}\right) \dfrac{(G_\gamma * u_0)_{x_1}(X_s)}{|\nabla(G_\gamma * u_0)(X_s)|}, & \dfrac{c(s)}{c} \dfrac{(G_\gamma * u_0)_{x_2}(X_s)}{|\nabla(G_\gamma * u_0)(X_s)|}
\end{bmatrix}
$$

$$
c(t) = \begin{cases}
0 & \text{if } t < S \quad \text{or} \quad |\nabla(G_\gamma * u_0)(x)| < d, \\
c & \text{if } t \geq S \quad \text{and } |\nabla(G_\gamma * u_0)(x)| \geq d,
\end{cases}
$$

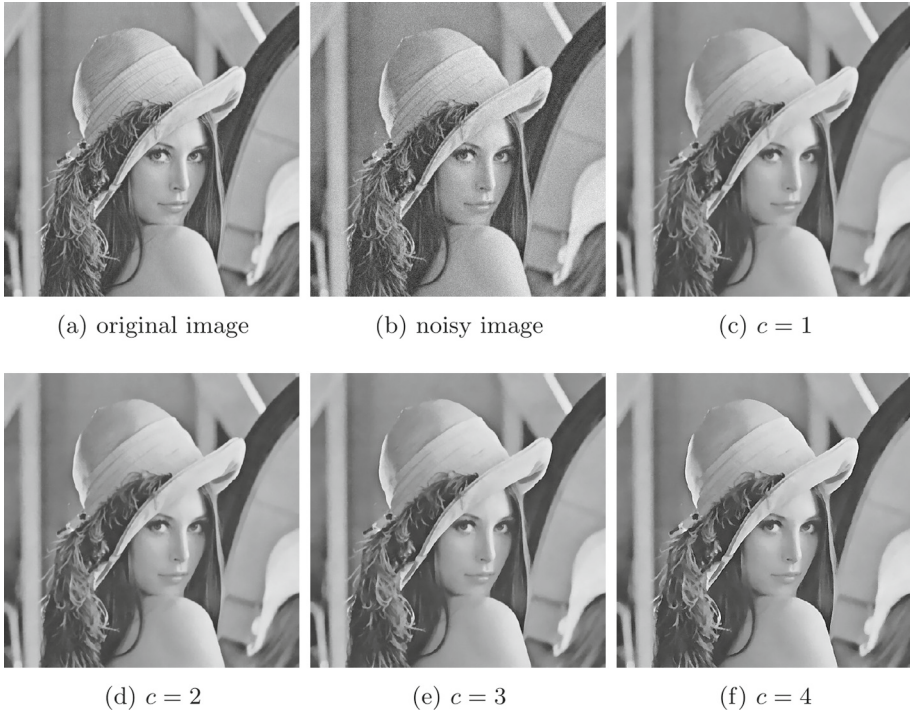$u_{x_i}(y) = \frac{\partial u}{\partial x_i}(y)$ and by $*$ we denote a convolution operation.

(a) original image   (b) noisy image   (c) $c = 1$

(d) $c = 2$ (e) $c = 3$ (f) $c = 4$

Fig. 1. Results of the reconstruction using [3] of grayscale Lenna image (512×512 pixels) for different values of the parameter c. Standard deviation of the noise is 10. The time of the reconstruction is $8\,\text{s}$ (CPU: $2 \times 2.5\,\text{GHz}$).

The Eq. (2) is a particular case of (1) for given specific functions σ and f. In this model we deblur in gradient direction from time T to S and smooth out in perpendicular to gradient direction from S to 0. To avoid false detections due to noise, u_0 is convolved with a Gaussian kernel G_γ (in practice 3×3 Gaussian mask). Parameter T defines the size of the neighbourhood used in the reconstruction procedure. The parameter d determines which pixels will be reconstructed with using smoothing model (only in gradient direction) and which with using enhancing model (in gradient and in perpendicular to gradient direction). The parameter c is responsible for effect of sharpening of edges (see Fig. 1).

If reconstructed pixel x belongs to the edge (the condition $|\nabla(G_\gamma * u_0)(x)| \geq d$ is true) the process X has values as in Fig. 2 and we will use pixels along edges (from time 0 to S) and pixels in gradient direction (from time S to T). The final value $u(x)$ is the sum of the pixel given by X_S multiplied by positive weight and pixels from time S to T multiplied by negative weights. If the condition $|\nabla(G_\gamma * u_0)(x)| \geq d$ is false then we explore only pixels along edges with positive weights and resign from sharpening effect.
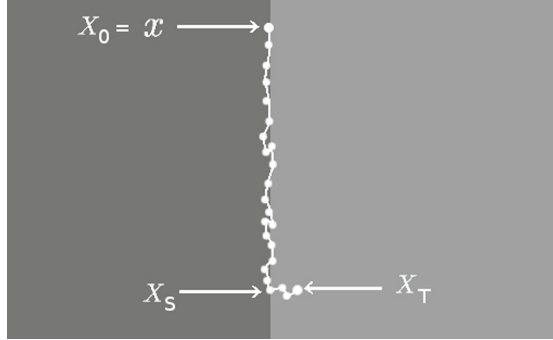
Fig. 2. Example of a trajectory of the process X defined by (2) which starts from the edge of the image.

4   Vector Valued Images

Now we concentrate on images with values in $\mathbf{R}^n$. A very common idea to restore vector-valued images is to use scalar diffusion on each channel of a noisy image. But one may quickly notice that this scheme is useless, since each image channel evolves independently with different smoothing geometries. To avoid this blending effect, the regularization process has to be driven in a common and coherent way for all vector image channels. In order to execute that we use Di Zenzo geometry [7,8]. Let $u : D \to \mathbf{R}^n$ be a vector valued image and $x \in D$ be fixed. Consider the function $F_x : V \to \mathbf{R}$, $F_x(v) = \left| \frac{\partial u}{\partial v}(x) \right|^2$, where $V = \{v \in \mathbf{R}^2; |v| = 1\}$. We are interested in finding the arguments $\theta_+(u,x), \theta_-(u,x)$ and corresponding values $\lambda_+(u,x) = F_x(\theta_+(u,x))$, $\lambda_-(u,x) = F_x(\theta_-(u,x))$ which maximize and minimize the function F_x, respectively. Note that F_x can be rewritten as $F_x(v) = F_x([v_1, v_2]^T) = v^T \mathbf{G}(x) v$. In the useful case of colour RGB images $\mathbf{G}$ is defined by the following [7]

$$
\mathbf{G}(x) = \begin{bmatrix} \sum\limits_{i=1}^{3} \left(\frac{\partial u_i}{\partial x_1}(x) \right)^2, & \sum\limits_{i=1}^{3} \frac{\partial u_i}{\partial x_1}(x) \frac{\partial u_i}{\partial x_2}(x) \\ \sum\limits_{i=1}^{3} \frac{\partial u_i}{\partial x_1}(x) \frac{\partial u_i}{\partial x_2}(x), & \sum\limits_{i=1}^{3} \left(\frac{\partial u_i}{\partial x_2}(x) \right)^2 \end{bmatrix}. \tag{3}
$$

The interesting point about $\mathbf{G}(x)$ is that its positive eigenvalues $\lambda_+(u,x)$, $\lambda_-(u,x)$ are the maximum and the minimum of F_x while the orthogonal eigenvectors $\theta_+(u,x)$ and $\theta_-(u,x)$ are the corresponding variation orientations. Three different choices of vector gradient norms $N(u,x)$ have been proposed in the literature: $\sqrt{\lambda_+(u,x)}$, $\sqrt{\lambda_+(u,x) - \lambda_-(u,x)}$ and $\sqrt{\lambda(u,x) + \lambda_-(u,x)}$. In this paper we have used $N(u,x) = \sqrt{\lambda_+(u,x)}$ as a natural extension of the scalar gradient norm viewed as the value of maximum variations.
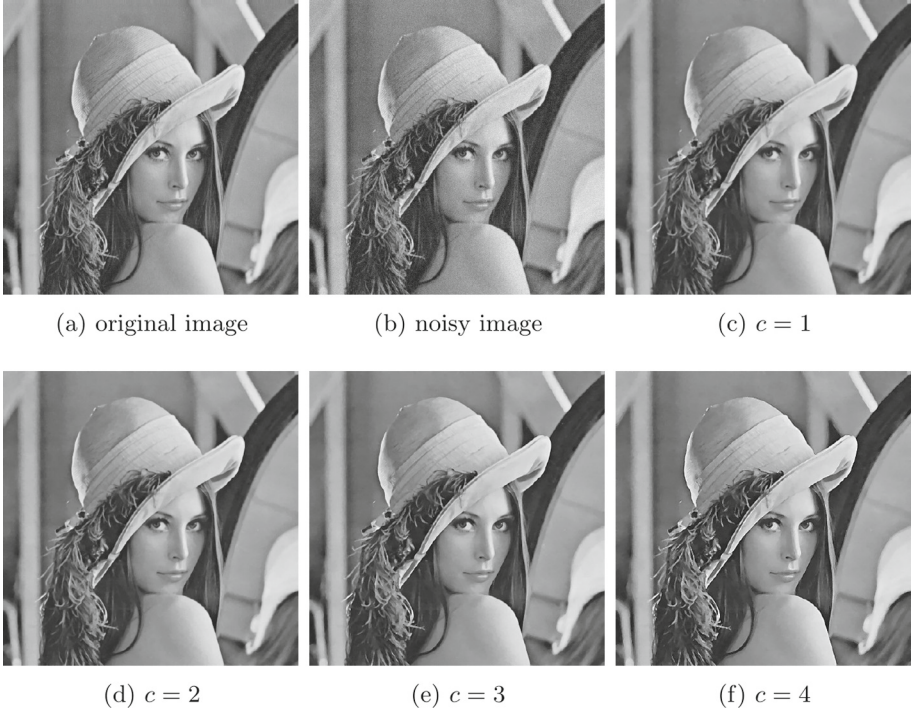
(a) original image   (b) noisy image   (c) $c = 1$

(d) $c = 2$ (e) $c = 3$ (f) $c = 4$

Fig. 3. Results of the reconstruction using Eq. (4) of RGB Lenna image (512×512 pixels) for different values of the parameter c. Standard deviation of the noise is 10 (independently added to all coordinates). The time of the reconstruction is 19 s (CPU: 2×2.5 GHz).

Replacing in Eq. (2) $|\nabla(u, x)|$ and $[u_{x_1}(x), u_{x_2}(x)]^T$ by $N(u, x)$ and $\theta_+(u, x), \theta_-(u, x)$ respectively, we obtain the following model for vector-valued images:

$$\begin{cases} X_t = x + \int_0^t \sigma(s, X_s)\, dW_s + K_t^{\overline{D}}, & t \in [0, T], \\ Y_t = u_0(X_S) + \int_t^T c(s)(Y_s - u_0(X_s))ds - \int_t^T Z_s\, dW_s, & t \in [0, T], \end{cases} \tag{4}$$

where $S < T$,

$$\sigma(s, X_s) = \left[\left(1 - \frac{c(s)}{c}\right)\theta_-(G_\gamma * u_0, X_s), \frac{c(s)}{c}\theta_+(G_\gamma * u_0, X_s) \right]$$

$$c(t) = \begin{cases} 0 & \text{if } t < S \quad \text{or} \quad N(G_\gamma * u_0, x) < d, \\ c & \text{if } t \geq S \quad \text{and} \quad N(G_\gamma * u_0, x) \geq d. \end{cases}$$

In Fig. 3 we can see similar as before results of the reconstruction using (4) for RGB Lenna image and different values of the parameter c. The parameter d depends on the noise of the image, and it can be set as the standard deviation of the noise.

5   Approximation

The following theorem allows us to transition from a continuous model to the discrete model.

Theorem 1. *Let $S < T$, $u_0 : \overline{D} \to \mathbf{R}^n$, $x \in \overline{D}$. Assume that $f(t, y) = c(t)(y - u_0(X_t))$, $c(t) = 0$ for $t < S$ and $\xi = u_0(X_S)$, where X is two-dimensional diffusion process with reflection with values in $\overline{D}$ and starting from x. If (Y, Z) is a solution to BSDE*

$$Y_t = \xi + \int_t^T f(s, Y_s)ds - \int_t^T Z_s\, dW_s, \quad t \in [0, T],$$

then

$$\lim_{m \to +\infty} Y_0^m = Y_0,$$

where

$$Y_0^m = \sum_{k=j}^{m-1} a_k \mathbf{E}\left[u_0(X_{t_k})\right], \tag{5}$$

and j is a index of discretisation, such that

$$0 = t_0 < t_1 < \cdots < t_j \le S < t_{j+1} < \cdots < t_m = T, \quad t_{i+1} - t_i = \frac{T}{m},$$

$$a_j = \prod_{r=j}^{m-1}\left(1 + \frac{c(t_r)T}{m}\right) - \frac{c(t_j)T}{m}, \tag{6}$$

$$a_k = -\frac{c(t_k)T}{m}\prod_{r=j}^{k-1}\left(1 + \frac{c(t_r)T}{m}\right), k = j+1, j+2, ..., m-1, \tag{7}$$

and by $\mathbf{E}$ we denote the expected value.

The coefficients a_k, $k = j, j+1, ..., m-1$ define the weights used in the process reconstruction and satisfy the useful condition $\sum_{k=j}^{m-1} a_k = 1$. The proof of the above theorem is analogous to the proof of Theorem 5.1 in [3]. Repeating arguments from the work [3] and using the above theorem we obtain the following discrete inverse formula for RGB images.

$$u(x) \approx \begin{cases} \dfrac{1}{N_0}\displaystyle\sum_{n_0=1}^{N_0} u_0(X_{t_j}^x(\omega_{n_0})) & \text{if } N(G_\gamma * u_0, x) < d, \\[2.5ex] \dfrac{1}{N_0}\sum_{n_0=1}^{N_0}\Big[a_j u_0\left(X_{t_j}^x(\omega_{n_0})\right) \\[1.5ex] \quad + \dfrac{1}{N_1}\sum_{n_1=1}^{N_1}\sum_{k=j+1}^{m-1} a_k u_0\left(X_{t_k}^{X_{t_j}^x(\omega_{n_0})}(\omega_{n_1})\right)\Big] & \text{if } N(G_\gamma * u_0, x) \ge d \end{cases} \tag{8}$$

where $X_t^a(\omega)$ is the value at time t of ω-trajectory of the process X starting from a point a.

6   Experimental Results

Measure of quality for our evaluation experiments regarding the new method and other classic methods: total variation [12] and non local means [6] for colour images are presented in Table 1. The results refer to RGB image *Lenna* corrupted (channels independently) with the Gaussian noise with standard deviation ρ. Noisy images have been reconstructed with vector analysis in RGB space. The maximum values of Peak Signal to Noise Ratio (in short PSNR) is given in the table. The analysis of PSNR measure shows that the new method performs better then PDE approach and is comparable to the method based on non-local similarity of patches.

Note that the type of high ISO noise generated by a typical digital camera sensor can be modelled as an additive white Gaussian distribution with standard deviation proportional to the value of the ISO [5]. In Fig. 4 we present the images taken with high ISO value and the result of the reconstruction using the new algorithm, non local means [6] and total variation [12]. Parameters of all methods were set to the default values as recommended by [3,6,12]. It should be noted that we have one unknown parameter: standard deviation of the noise. This parameter is determined from the background of the image and set to 5. Looking (carefully) at the images you can observe that the image created by the new method is visually more sharp and therefore all the details are more visible (see Fig. 4).

Table 1. Lenna (PSNR)

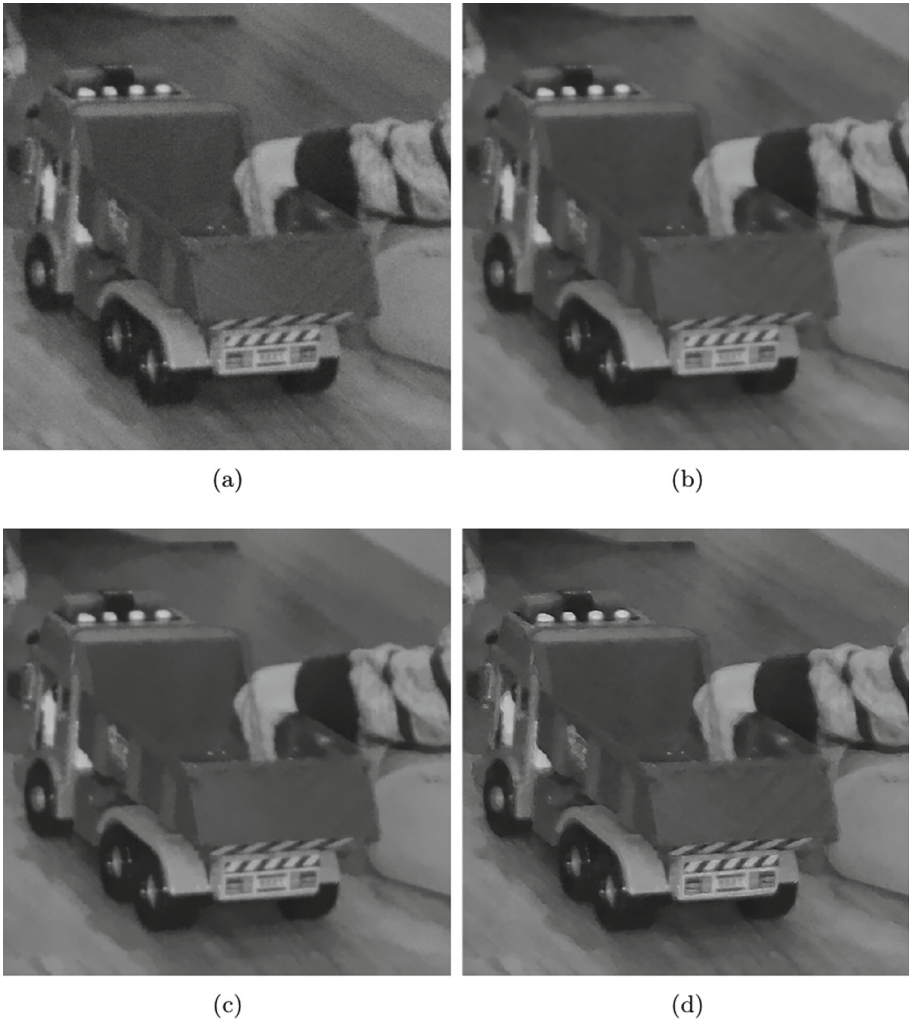| ρ	Non local means [6]	New method ($c = 1.5$)	Total variation [12]
10	**34.7758**	33.4106	33.2549
15	**33.1034**	32.2666	31.7692
20	**31.9212**	31.3412	30.7777
25	**30.8765**	30.6641	29.9996
30	**30.1963**	30.0058	29.3863
35	**29.5205**	29.4037	28.8702
40	28.8700	**28.8767**	28.4306
45	28.3743	**28.3912**	28.0531
50	27.7861	**27.9544**	27.7119
55	27.3753	**27.5024**	27.4102

Fig. 4. (a) Input high ISO image (b) Result of reconstruction using total variation [12] (c) Result of reconstruction using non local means [6] (d) Result of reconstruction using the new method

7   Conclusion

In this paper we proposed a new method of colour image denoising. The obtained results demonstrate that proposed approach can be used successfully to reconstruction of high ISO images and can provide a good alternative to other methods.

References

1. Borkowski, D.: Chromaticity denoising using solution to the skorokhod problem. In: Image Processing Based on Partial Differential Equations, pp. 149–161 (2007)
2. Borkowski, D.: Smoothing, enhancing filters in terms of backward stochastic differential equations. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, pp. 233–240. Springer, Heidelberg (2010)
3. Borkowski, D.: Forward and backward filtering based on backward stochastic differential equations. Inverse Probl. Imaging **10**(2), 305–325 (2016)
4. Borkowski, D., Jańczak-Borkowska, K.: Application of backward stochastic differential equations to reconstruction of vector-valued images. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012, pp. 38–47. Springer, Heidelberg (2012)
5. Bovik, A.C.: Handbook of Image and Video Processing. Academic Press, Amsterdam (2010)
6. Buades, A., Coll, B., Morel, J.M.: Non-local means denoising. Image Process. On Line **1**, 208–212 (2011)
7. Deriche, R., Tschumperlé, D.: Diffusion PDE's on vector-valued images: local approach and geometric viewpoint. IEEE Sig. Process. Mag. **19**(5), 16–25 (2002)
8. Di Zenzo, S.: A note on the gradient of a multi-image. Comput. Vis. Graph. Image Process. **33**(1), 116–125 (1986)
9. Duffie, D., Epstein, L.G.: Asset pricing with stochastic differential utility. Rev. Fin. Stud. **5**, 411–436 (1992)
10. Duffie, D., Epstein, L.G.: Stochastic differential utility. Econometrica **60**, 353–394 (1992)
11. El Karoui, N., Peng, S., Quenez, M.C.: Backward stochastic differential equations in finance. Math. Fin. **7**(1), 1–71 (1997)
12. Getreuer, P.: Rudin-osher-fatemi total variation denoising using split bregman. Image Process. On Line **2**, 74–95 (2012)
13. Hamadene, S., Lepeltier, J.P.: Zero-sum stochastic differential games and backward equations. Syst. Control Lett. **24**(4), 259–263 (1995)
14. Pardoux, E., Peng, S.: Adapted solution of a backward stochastic differential equation. Syst. Control Lett. **14**(1), 55–61 (1990)
15. Pardoux, E., Tang, S.: Forward-backward stochastic differential equations and quasilinear parabolic pdes. Probab. Theory Relat. Fields **114**(2), 123–150 (1999)