Non-local Means for Stereo Image Denoising Using Structural Similarity

Monagi H. Alkinani and Mahmoud R. El-Sakka(✉)

Computer Science Department, University of Western Ontario,
London, ON N6A 5B7, Canada
{malkinan,melsakka}@uwo.ca

Abstract. We present a novel stereo image denoising algorithm. Our algorithm takes as an input a pair of noisy images of an object captured form two different directions. We use the structural similarity index as a similarity metric for identifying locations of similar patches in the input images. We adapt the Non-Local Means algorithm for denoising collected patches from the input images. We validate our algorithm on various stereo images at various noise levels. Experimental results show that the denoising performance of our algorithm is better than the original Non-Local Means and Stereo-MSE methods at low noise level ($\sigma \leqslant 20$).

Keywords: Non-local means · Patch-based image filtering · Stereo imaging · Structural similarity index · Additive noise reduction · Disparity map

1 Introduction

Digital images are captured using sensors during the data acquisition phase, where they are often contaminated with an undesired random noise. Such noise can also be produced during image transmission or image compression. Additive noise is generally modelled as:

$$\upsilon(x) = u(x) + n(x)_d, x \in \Omega \tag{1}$$

where $\upsilon(x)$ is the noisy component of the image, $u(x)$ is the true image, $n(x)_d$ is the random additive noise, and Ω denotes the set of all pixels in the image. In particular, if $n(x)_d$ is a Gaussian random process, then the noise is identified as a Gaussian noise. The noise level in digital images vary from being almost imperceptible to being very noticeable. Image denoising schemes attempt to produce a new image that has less noise, i.e., closer to the noise-free image $u(x)$.

Denoising techniques can be grouped into two main approaches: pixel-based image filtering and patch-based image filtering. A pixel-based image filtering scheme is mainly a proximity operation used for manipulating one pixel at a time based on its spatial neighbouring pixels. Such methods include low-pass filtering using Gaussian filter [1], Yaroslavsky filter [2], bilateral filter [3], total variation filter [4], and anisotropic diffusion filter [5]. On the other hand, in patch-based image filtering, the noisy image is divided into patches, or "blocks",

M. Kamel and A. Campilho (Eds.): ICIAR 2015, LNCS 9164, pp. 51–59, 2015.
DOI: 10.1007/978-3-319-20801-5_6

which are then manipulated separately in order to provide an estimate of the true pixel values based on similar patches located within a searching window. Such methods include Non-Local Means [6], patch-based PCA [7], K-SVD [8], and BM3D [9]. Patch-based image filtering approach utilizes the redundancy and the similarity among the various parts of the input image.

Non-Local Means (NL-Means) is a patch-based image filter proposed by Buades *et al.* [6] as a modification to the pixel-based bilateral filter. Like the bilateral filter, it blurs the homogeneous areas and preserves edges.

As a new application for NL-Means filter, we would like to adapt it for denoising stereo images in order to improve the extracted depth information coming from noisy stereo images. A stereo image uses two or more images generated from cameras at different locations. By computing the differences between the images, the depth information can be extracted. Noisy stereo images would give disappointing results when they are used for extracting depth information. In this work, NL-Means is utilized for denoising stereo images. Our proposed method extends the searching window to search the two images when seeking similar patches.

Using multi-view images for noise reduction has a unique advantage over using only one-view image. In multi-view images, a pixel in one image is estimated based on the corresponding pixels from all other images. This approach is popular in video denoising where multi-frames are used for noise reduction [10,11]. Recently, great progress has been made to break the limits of using one input image when denoising 3D images. Zhang *et al.* extended the idea of using patch-based PCA for denoising single image to multi-view images [12]. While patch-based PCA collects similar patches locally and globally from single image before applying the PCA algorithm, Zhang *et al.* algorithm collects similar patches from multiple images. Heo *et al.* use Maximum A Posteriori-Markov Random Field (MAP-MRF) as a model for energy minimization in order to find the disparity maps from stereo image [13]. In order to find the disparity maps, they proposed an algorithm that consists of two terms: the first term is the restored intensity difference, and the second term is the dissimilarity of support pixel distribution. They adapted NL-Means algorithm for the restoration of intensity values of the first term. They extended the NL-Means algorithm for denoising stereo images by grouping similar patches by using MSE from two similar windows in left and right images, we called this method Stereo-MSE.

The rest of the paper is organized as follow. Section 2 introduces the NL-Means filter and its mathematical formulation. Section 3 describes our proposed method. In Sect. 4, we compare the performance of the proposed method with other denoising filters. Section 5 offers concluding comments, and future works.

2 NL-Means Algorithm

The NL-Means filter divides the input image into sub-images and then filters each sub-image separately in a technique that is referred to as being patch-based. Each sub-image contains several patches. As in the bilateral filter, the similarity in NL-Means filter is assessed based on two measurements: (1) the Euclidean distance between the centres of the patches, and (2) the luminance distance between the patches. In contrast to the bilateral filter, NL-Means filter uses patches from a searching window instead of using single neighbouring pixels when assigning weights and averaging. This is why it is called a non-local method. Patches with similar grey levels are assigned larger weights when averaging. Similar to bilateral filter, NL-Means filter preserves edges regardless of their directions.

Equation 2 is used to estimate a pixel i using NL-Means filter,

$$NLMeans\,[v]_i = \sum_{j \in I} \omega(i,j)\,[v]_j \tag{2}$$

where $[v]_i$ and $[v]_j$ are pixels intensities at location i and j, respectively, and $\omega(i,j)$ is a similarity measure between pixels i and j. The similarity weight, $\omega(i,j)$, satisfies the condition $0 \leq \omega(i,j) \leq 1$ and $\sum_j \omega(i,j) = 1$. It depends on the grey level similarity and the Euclidean distance between vectors $N\,[v]_i$ and $N\,[v]_j$, where $N\,[v]_k$ denotes a square neighbourhood of fixed size and centred at a pixel k. The weights are described as,

$$\omega(i,j) = \frac{1}{Z(i)} e^{-\frac{\|(N[v]_i) - \left(N[v]_j\right)\|^2}{h^2}} \tag{3}$$

where $Z(i)$ is a normalization factor and h is a filtering parameter set depending on the noise level.

The level of noise determines the sizes needed for the searching window and patches. For a robust comparison between patches, the size of the patches increases when the noise level is high. Accordingly, the value of the filtering parameter h increases as the size of the patch is increased. Meanwhile, the size of the searching window must be increased in order to find more similar patches.

3 Proposed Algorithm

In this section, we describe a new algorithm for solving the problem of denoising stereo images. The novelty of this algorithm is the use of the NL-Means algorithm to denoise multi-view images. We increase the number of similar patches by grouping similar patches from left and right images of a stereo image. Figure 1 shows the way of collecting similar patches.

3.1 Algorithm Outline

Our algorithm is illustrated in Fig. 2. The left stereo image is processed in a raster scan. At each pixel, the following procedure is performed:

1. Obtain from the left image a fixed-size square patch "reference patch" $N\,[vl]_k$ centred at location k.
2. Use the structural similarity (SSIM) index [14] to find from the right image the best patch $N\,[vr]_q$ centred at location q that is similar to the reference patch and identify its window location.

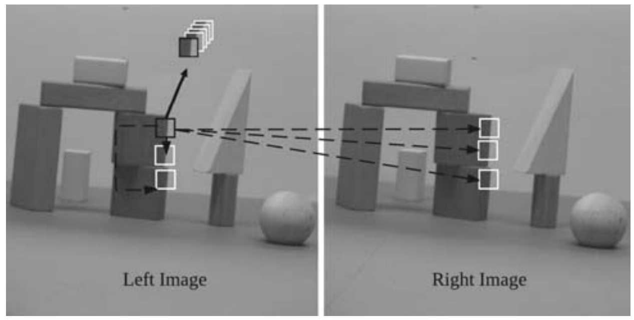

Fig. 1. Collecting similar patches from a stereo image: the patch with a black border is the reference patch, and the patches with white borders are similar patches

3. Collect patches from the two windows (using MSE) and assign weights ω to each patch. Similar patches to the reference patch are assigned high weights. The weights are assigned as described in Eq. 3.
4. Calculate the weighted average of patches, in order to estimate the true pixel of the left image. The estimated value $NLMeans\,[vl]_i$, for a pixel i located in the left image, is computed as:

$$NLMeans\,[vl]_i = \sum_{j \in I} \omega(i,j)\,[v]_j \tag{4}$$

   where $[vl]_i$ and $[v]_j$ are pixel intensities at location i in the left image and j from the left or right image, and $\omega(i,j)$ is a similarity weight between pixels i and j.

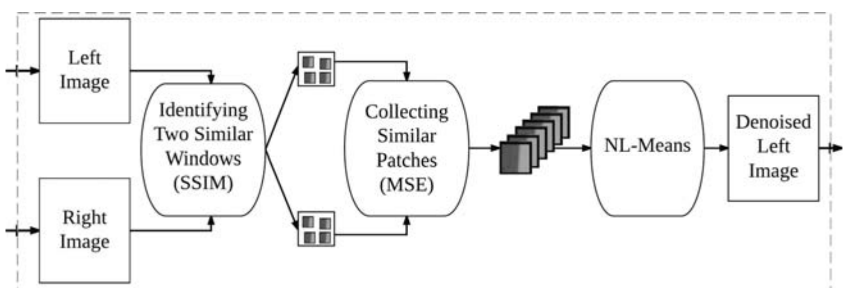

Fig. 2. A block digram of the proposed denoising method for stereo image denoising

3.2 Structural Similarity Index

Patch-based denoising methods achieve better results when there are enough similar patches that are accurately grouped before starting the actual denoising process. Choosing an accurate similarity metric would improve the whole denoising process. The main contribution of our method is to use the structural similarity index as a similarity metric for extending the search area, which makes our algorithm groups better similar patches from left and right images.

SSIM index is a metric for measuring the similarity between two images. Unlike the traditional approaches, e.g., peak signal-to-noise ratio (PSNR) and mean squared error (MSE), SSIM has been proven to be consistent with human perception. SSIM considers image degradation as perceived change in structural information, traditional approaches estimate perceived errors in image data. The SSIM metric between two patches of size $n \times n$ is calculated as:

$$SSIM(N\left[vl\right]_k, N\left[vr\right]_q) = \frac{\left(2\mu_L\mu_R + C_1\right)\left(2\sigma_{LR} + C_2\right)}{\left(\mu_L^2 + \mu_R^2 + C_1\right)\left(\sigma_L^2 + \sigma_R^2 + C_2\right)} \tag{5}$$

where $N\left[vl\right]_k$ is a reference patch from the left image, $N\left[vr\right]_q$ is a corresponding patch from the right image, μ_L and μ_R are the mean of the reference and corresponding patches, respectively. σ_L^2 and σ_R^2 are the variance of the reference and corresponding patches, respectively. σ_{LR} is the covariance between the reference and the corresponding patches. C_1 and C_2 are constants used to avoid instability.

4 Experimental Results

The objective of this section is to experimentally study the performance of the proposed method at various noise levels. The complexity of our algorithm is linear with respect to the size of stereo input image. We use a fixed 5×5 patch size and a fixed 11×11 searching window size. Four stereo images are used to perform this experiment. The four images are grey-scale images, they are shown in Fig. 3. MatLab is used for this experiment. The computer's processor is Intel® Core^{TM} i7 (2.5 GHz). In Subsects. 4.1 and 4.2, the methods are evaluated both quantitatively and qualitatively, respectively.

4.1 Quantitative Evaluation

Image Similarity Metrics. Two image similarity metrics are used for objective comparison between the results: (1) SSIM, and (2) peak signal-to-noise ratio (PSNR). The best result for SSIM is 1, while the PSNR has good result when its value is high. Equations 5 and 6 show the formulas for these two quality metrics. The peak signal-to-noise ratio is defined as:

$$PSNR = 10\log\left(\frac{\left(2^n - 1\right)^2}{MSE}\right) \tag{6}$$

where n is an integer number representing the number of bits per pixel. When $n = 8$, i.e., in case of grey-scale images.

It is worth mentioning that a study conducted by Horé *et al.* [15] has revealed that SSIM is less sensitive to additive noise than PSNR. They used F-score test to compare between SSIM and PSNR performances with additive Gaussian white noise.

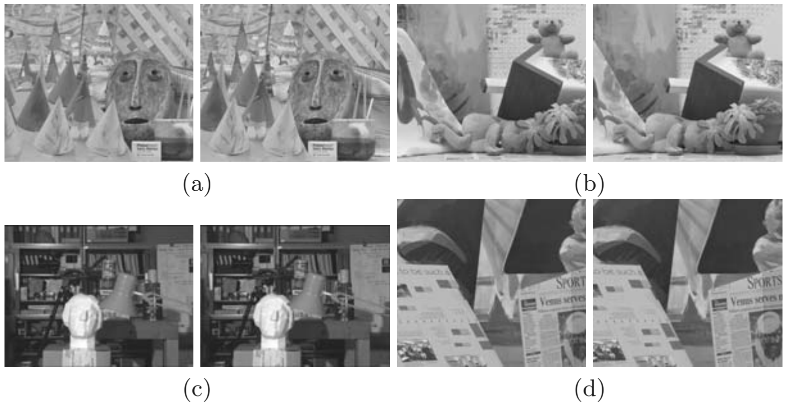

Fig. 3. The four used images in the experiment: (a) *cones* images 450×375, (b) *teddy* image 450×375, (c) *tsukuba* image 384×288, and (d) *venus* image 434×383.

Experimental Results. The experimental results of our proposed method are shown in Table 1, which compares the performance of our method with two other denoising methods: the original NL-Means [6] and Stereo-MSE [13]. The highest values of SSIM are highlighted by a bold font with a wavy under-bar, while the highest values of PSNR are highlighted with a bold font. The results in Table 1 are computed by measuring the differences between the true original images and the denoised images. At low noise level ($\sigma \leqslant 20$) our method performs better than the original NL-Means and Stereo-MSE (from both SSIM and PSNR point of views).

Table 1. The performance of the denoising algorithms at various noise levels (σ).

	σ	$\sigma = 10$		$\sigma = 20$		$\sigma = 40$		$\sigma = 60$	
	Method	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
Cones	Noisy	0.5726	25.067	0.3538	21.138	0.1651	15.914	0.0973	12.914
	NLMeans	0.7208	26.648	0.6111	25.254	0.4385	23.083	0.3462	**21.768**
	Stereo-MSE	0.7204	26.620	0.6168	25.321	0.4356	23.052	0.3162	21.198
	Our Meth	**0.7273**	**26.754**	**0.6397**	**25.555**	**0.4758**	**23.395**	**0.3553**	21.490
Teddy	Noisy	0.5430	25.681	0.3170	21.42	0.1472	16.160	0.0870	13.191
	NLMeans	0.7763	27.886	0.6686	26.471	0.4958	24.155	**0.4000**	**22.62**
	Stereo-MSE	0.7813	27.927	0.6670	26.524	0.4586	23.884	0.3323	21.805
	Our Meth	**0.7924**	**28.070**	**0.7058**	**26.870**	**0.5284**	**24.487**	0.3970	22.255
Tsukuby	Noisy	0.5884	25.635	0.3673	21.670	0.1830	16.792	0.1067	13.823
	NLMeans	0.8009	27.658	0.7028	26.180	**0.5374**	23.654	**0.4289**	**21.743**
	Stereo-MSE	0.8105	27.862	0.7051	26.402	0.5084	23.693	0.3727	21.559
	Our Meth	**0.8216**	**28.023**	**0.7308**	**26.571**	0.5115	**24.123**	0.4043	21.726
Venus	Noisy	0.4941	24.901	0.2850	21.138	0.1401	16.235	0.0845	13.275
	NLMeans	0.7572	26.779	0.6523	25.681	0.5069	23.745	**0.4119**	**22.165**
	Stereo-MSE	0.7575	26.783	0.6396	25.658	0.4523	23.589	0.3324	21.651
	Our Meth	**0.7707**	**26.929**	**0.6725**	**25.907**	**0.5424**	**23.789**	0.3913	22.153

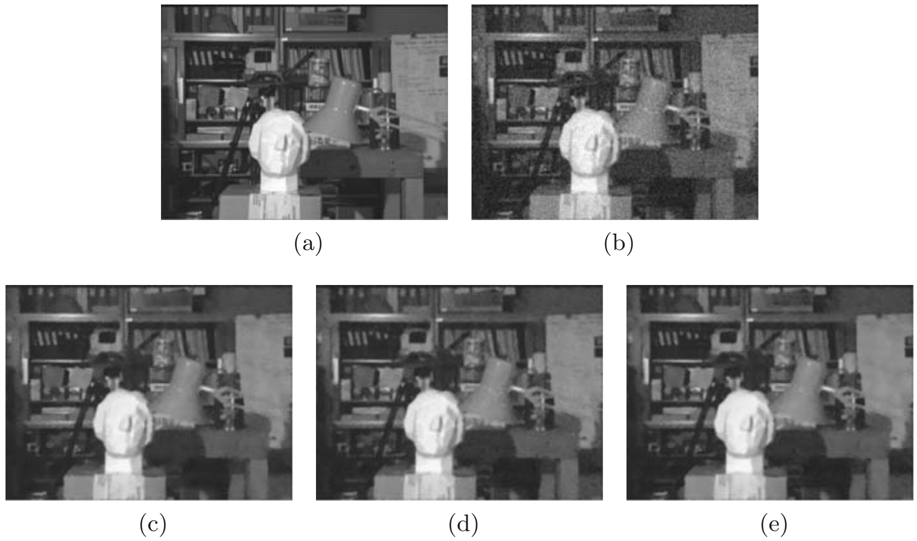

Fig. 4. The results of the denoising methods when denoising Tsukuba image at noise levels ($\sigma = 20$): (a) $Tsukuba$ image 384 × 288, (b) AWGN image, ($\sigma = 20$), (c) NL-Means, (d) Stereo-MSE, and (e) Our method

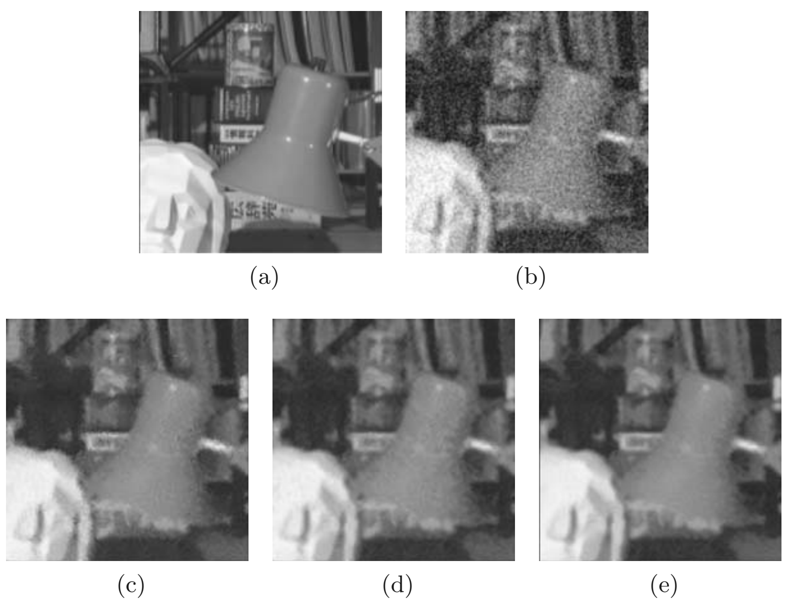

Fig. 5. Zoomed images of the denoised $Tsukuba$ images shown in Fig. 4: (a) $Tsukuba$ image, (b) AWGN image ($\sigma = 20$), (c) NL-Means, (d) Stereo-MSE and (e) Our method

Stereo-MSE and our algorithm are slower than the original NL-Means, as they search both images, not just a single image like the original NL-Means.

4.2 Qualitative Evaluation

The evaluation in this section is subjective, where the quality of the denoised images is addressed via the visual perception. Denoised Tsukuba images with AWGN ($\sigma = 20$) are chosen to perform this evaluation. The results of denoising `Tsukuba`'s image are shown in Fig. 4.

Figure 4e shows that our method achieved the best results. Our method preserves sharp edges; i.e., the books in the background of `Tsukuba` image. Homogeneous regions are smoothed properly by our method; i.e., head and lamp in the `Tsukuba` image. Our method does not restored clearly words with small font size written on the board that shown in the `Tsukuba` image. A zoomed version of Fig. 4 is shown in Fig. 5.

5 Conclusion and Future Work

In this paper, we looked at stereo image as a multi-view image and sought to restore left image by using SSIM and NL-Means approaches. Empirical results show that our method achieved better denoising than the original NL-Means and Stereo-MSE methods, at low noise level ($\sigma \leqslant 20$). Stereo-MSE and our method are slower than the original NL-Means. We believe that our work opens several interesting doors for future work. First, our current method does not consider denoising right image at the same time when it denoising left image. We believe that denoising left and right image at the same time would produce two denoised image in shortest time. Second, our algorithm does not use SSIM for assigning the weights between similar patches. Since SSIM combats the traditional similarity metrics, we believe that using SSIM as a similarity metric for assigning weights would help to improve our algorithm. Last, the speed of our algorithm could be reduced when the interesting search region in the right image is reduced.

References

1. Lindenbaum, M., Fischer, M., Bruckstein, A.: On Gabor's contribution to image enhancement. Pattern Recogn. **27**(1), 1–8 (1994)
2. Yaroslavsky, L.P.: Digital Picture Processing: An Introduction. Springer, Heidelberg (1985). ISBN 3-540-11934-5
3. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: 1998 IEEE International Conference on Computer Vision, pp. 839–846, Bombay, India (1998)
4. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. In: 11th Annual international Conference of the Center for Nonlinear Studies on Experimental Mathematics, vol. 60(1–4), pp. 259–268. Elsevier North-Holland Inc., Amsterdam (1992)

5. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Analysis Mach. Intel. **12**, 629–639 (1990)
6. Buades, A., Coll, B., Morel, J.: A non-local algorithm for image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 2, pp. 60–65 (2005)
7. Deledalle, C., Salmon, J., Dalalyan, A.: Image denoising with patch based PCA: local versus global. In: Hoey, J., McKenna, S., Trucco, E. (eds.) Proceedings of the British Machine Vision Conference, pp. 25.1–25.10. BMVA Press, Durham (2011)
8. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**(11), 4311–4322 (2006)
9. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**, 2080–2095 (2007)
10. Bennett, E.P., McMillan, L.: Video enhancement using per-pixel virtual exposures. ACM Trans. Graph. **24**(3), 845–852 (2005)
11. Danielyan, A., Foi, A., Katkovnik, V., Egiazarian, K.: Image and video super-resolution via spatially adaptive blockmatching filtering. In: SPIE Electronic Imaging, 2008, no. 6812–07, San Jose, California, USA (2008)
12. Zhang, L., Vaddadi, S., Jin, H., Nayar, S.K.: Multiple view image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 1542–1549 (2009)
13. Heo, Y., Lee, K., Lee, S.: Simultaneous depth reconstruction and restoration of noisy stereo images using non-local pixel distribution. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8 (2007)
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)
15. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: 20th International Conference on Pattern Recognition (ICPR), pp. 2366–2369 (2010)