Weighted MSE Based Spatially Adaptive BM3D

Mykola Ponomarenko, Marina Pismenskova, and Karen Egiazarian
Laboratory of Signal Processing,
Tampere University of Technology,
P.O. Box 553, FI-33101 Tampere, Finland
mykola.ponomarenko@tut.fi, marina.pismenskova@tut.fi, karen.eguiazarian@tut.fi,

***Abstract*— Weighted MSE (wMSE), recently introduced modification of MSE, is an image quality metric used to estimate visual quality of filtered images. It provides better than MSE correspondence to a human perception in consideration of distortions introduced by image filters. In this paper, wMSE is used both as a criterion to evaluate filtering efficiency of the modification of BM3D filter with spatially varying parameters, as well as to train a specially designed neural network to predict filters' parameters. Extensive analysis on three image datasets demonstrates that the proposed modification of BM3D provides lower values of wMSE than those of BM3D, both effectively suppressing noise in homogeneous regions as well as preserving fine details and texture.**

Keywords—image denoising, image visual quality assessment, neural networks, BM3D

I. Introduction

Image denoising is an area of intense research with a large number of new image denoising methods proposed every year. For performance evaluation of an image denoising method usually full-reference image quality metrics are used, among them mean square error (MSE) and, its derivative, peak signal to noise ratio (PSNR) are the most widespread. At the same time, other image quality metrics, coping better with peculiarities of human visual system, become more and more popular [1-3]. However, these metrics still lack high enough correlation with a human perception [2] to assure their efficient usage in digital image processing.

The outputs of most full-reference metrics produce single values to characterize a difference between distorted and reference images. Image homogeneous regions, for which noise suppression is the most efficient, often has a larger contribution to the metrics' value than regions with edges and fine details. Because of this, an effective denoising of homogeneous regions, resulting in a better quality metric value, can compensate a slight decrease of the metrics' value due to over-smoothing of fine details and texture (usually occupying much less space in images comparing to homogeneous regions). However, as it was demonstrated in [4], details and texture preservation in noisy images is often more important for a human observer than effective noise suppression in homogeneous regions. This is also valid for many tasks of automatic image interpretation and object recognition, where a quality of solving these tasks strongly depends on a level of preservation of fine details and texture in filtered images.

In [4], two image quality metrics, namely, a weighted modification of MSE metric (wMSE) and the corresponding modification of PSNR metric, wPSNR were proposed. In order to compute an wMSE value, three images shall be used: original, noisy and filtered. Two pointwise distances are computed: between noisy and original (D_1), and between filtered and original images (D_2). Larger weights (by 5 times, found empirically in [4]) are assigned to those pixels for which values of D_2 are larger than values of D_1. This reflects the observation that for a human perception distortions introduced by filtering have larger weights than distortions eliminated by the filtering.

Due to this, it is important to consider the following problem. How to modify a filtering method taking into account wMSE or other metrics (e.g. weighted modifications of metrics PSNR-HVS [5] and PSNR-HVS-M [6], considered in [4])? This is not a trivial task, since one needs to know a reference image in order to compute such a metric, and a reference image is unavailable in practice. Thus, the only way to resolve this is to predict a value of such a metric inside the filtering method. Current paper is devoted to study the above problem.

In this paper, we modify one of the state-of-the-art in image denoising, the BM3D filter [7]. For BM3D, it is possible to control a degree of noise suppression by changing the threshold value at the hard-thresholding stage. A good noise suppression and details preservation is provided by using a standard filtering profile that uses the threshold value equal to 2.7σ [7] (noise is zero-mean Gaussian noise with variance σ^2). Let us denote this profile by *S-BM3D*. Possibly better details preservation but worse noise suppression can be achieved if one uses lower threshold value, e.g. equal to 2σ. Let us denote this profile by *P-BM3D*. Main idea of the proposed approach is to design a neural network to decide effectively and pixelwise between outputs of S-BM3D and P-BM3D. For the training of a network, we use maps of values of wMSE for outputs of S-BM3D and P-BM3D. As a neural network inputs, several pixelwise or patch-wise features of noisy image are calculated. The trained network shall produce relative weights for outputs of S-BM3D and P-BM3D. Thus, the output of the proposed spatially adaptive BM3D (called *A-BM3D*) will be a linear combination of outputs of different profiles of BM3D (namely, S-BM3D and P-BM3D).

The paper is organized as follows. In Section II the description of the proposed filter is provided. In Section III, an extensive analysis of A-BM3D in a comparison with two underlying profiles of BM3D is carried out. Finally, the conclusion follow.

ISBN 978-0-9928626-7-1 © EURASIP 2017

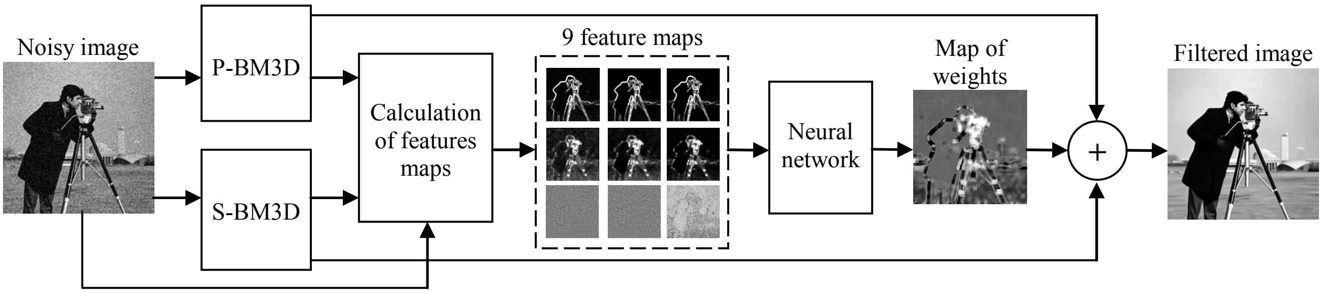


Fig. 1. Flow chart of the proposed A-BM3D filter

II. Description of the proposed filter

A. Main idea

Flow chart of the proposed A-BM3D filter is given in Fig. 1. Output of the filter for a given pixel is computed as a pixelwise weighted sum of the outputs of P-BM3D and S-BM3D:

$$L(i,j) = P(i,j)w(i,j) + S(i,j)(1-w(i,j)). \quad (1)$$

Here L(i,j) is an output of the A-BM3D filter for a pixel at location (i,j), P(i,j) is the output of P-BM3D filter, S(i,j) is the output of S-BM3D filter, w(i,j) is the weight, calculated by neural network, i=1,...,N and j=1,...,M are indexes of the image pixels, N x M is the size of the image.

The neural network with 9 inputs and 1 output was used in this work. Nine feature maps are uses as the input to the neural network. They are calculated using noisy image and outputs of P-BM3D and S-BM3D filters. Details of their construction will be given below in subsection II.C.

Note that the value of σ is required to calculate feature maps as well as to calculate outputs of P-BM3D and S-BM3D. This value is pre-estimated by one of blind noise parameters estimation methods (e.g. [8]) if not known in advance.

B. Maps of weights for learning the neural network

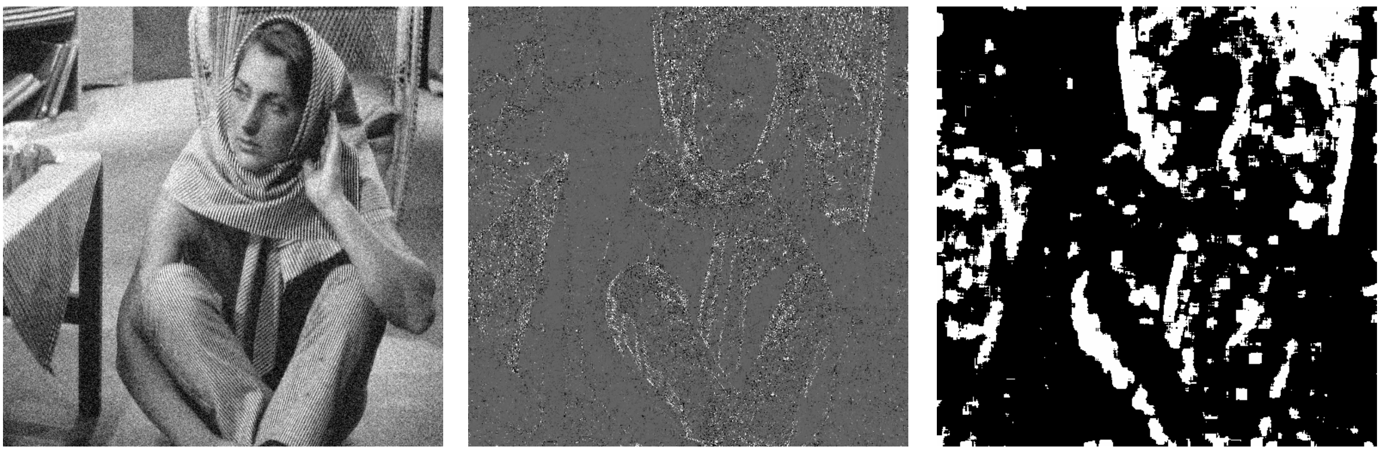
In this subsection we explain in details the goal of the designed neural network and the training procedure.

The aim of the designed neural network is to effectively switch (alternate) between outputs of P-BM3D and S-BM3D in order to provide the lowest possible value of the wMSE metric. To do this, we have used Δ maps, where Δ(i,j) is defined as:

$$\Delta(i,j) = \begin{cases} D_2(i,j), & D_2(i,j) \le D_1(i,j) \\ 5D_2(i,j), & D_2(i,j) > D_1(i,j) \end{cases}. \quad (2)$$

Fig. 2,b shows an example of the difference between Δ maps calculated for outputs of S-BM3D and P-BM3D (for the reference image Barbara and noisy image presented in Fig. 2,a). Bright pixels in Fig 2,a indicate that P-BM3D provides lower values of Δ than S-BM3D, and vice versa.

Since the image in Fig. 2,b is noisy and may be not so suitable for training, we have pre-filtered it by the mean filter with the window size 13. A binarized result after filtering is presented in Fig. 2,c, where white pixels indicate that P-BM3D provides lower Δ values and black pixels indicate that S-BM3D provides better quality.



a b c

Fig. 2. a) Noisy test image Barbara, σ^2=400, b) Difference between values of Δ for S-BM3D and P-BM3D, c) Smoothed in 13x13 window and binarized difference between values of Δ for S-BM3D and P-BM3D

Looking at Fig. 2,c, one may think that P-BM3D provides better visual quality than S-BM3D for image regions with high local energy (edges, textures, contrast fine details). However, it is not always so. For example, the table's leg in Fig. 2 is an object with a high local energy (along the border of the leg). Nevertheless, there are black pixels in Fig. 2,c corresponding to the table's leg, i.e. S-BM3D provides better visual quality for that region. This can be explained by a high level of self-similarity of such image regions and by the ability of BM3D to extract and preserve information from regions with high self-similarity.

It appears to be more reasonable (found empirically) to assume that P-BM3D provides better denoising for regions where high local energy is combined with a low self-similarity. This assumption will be taken into account in the selection of image features for inputs of the neural network.

C. Feature maps

As inputs of the neural network, the following 9 features calculated for each image pixel with indexes (i,j) have been selected:

#1: In the noisy image local variance σ^2_{loc} is calculated for patch of 9x9 pixels (top-left corner pixel of the patch has coordinates i-4, j-4). The feature #1 is calculated as σ_{loc}/ σ;

#2: This feature is calculated similarly to the feature #1, only using an output of P-BM3D instead of the noisy image;

#3: This feature is calculated similarly to the feature #1, only using an output of S-BM3D instead of the noisy image. Feature map #3 for the image from Fig. 2,a is shown in Fig. 3,a;

#4: Let us define the patch of 9x9 pixels of the noisy image, denoted by **A**. For the patch **A,** a patch **B** is searched in the area (i-15,...,i+15, j-15,...,j+15) to minimize the value of root mean square error (RMSE) between **A** and **B**. The feature #4 is calculated as RMSE(A,B)/ σ.

#5: This feature is calculated similarly to the feature #4, only using an output of P-BM3D instead of the noisy image;

#6: This feature is calculated similarly to the feature #4, only using an output of S-BM3D. Feature map #6 for the image in Fig. 2,a is shown in the Fig. 3,b;

#7: This feature is calculated as a pixelwise difference between the output of P-BM3D and the noisy image. The difference is normalized (divided by σ).

#8: The feature is calculated by a pixelwise difference between output of S-BM3D and the noisy image. The difference is normalized (divided) by σ.

#9: The feature is calculated as a pixelwise difference between absolute values of features #7 and #8.

Thus, among 9 feature maps, the first 3 are calculated using local variances, next 3 are calculated using dissimilarity maps, and the last 3 are calculated based on pixelwise differences between noisy and filtered images.

a b

Fig. 3. a) Feature map #3 for image from Fig. 2a, b) Feature map #6 for image from Fig. 2a.

As a result, for each pixel of the given noisy image, values of 9 features are calculated. If the reference image is available, then it is possible to calculate also the map of weights similar to the map shown in Fig 2,c. In the aggregate, values of these features and the values of calculated weights (for all pixels or for a part of image pixels) are used to train the neural network.

Note that the feature map #3 and feature map #6, presented in Fig. 3, used together, allow neural network to distinguish the table leg (considered in example above) from other objects with high local energy.

D. Configuration and learning the neural network

We have considered different configurations of the neural network. Best result has been achieved by applying the perceptron, consisting of 32 neurons at the input layer (activation function *tansig*), 16 neurons at the hidden layer (activation function *tansig*) and 1 neuron at the output layer (linear activation function).

As the training data, values of 9 features maps (inputs) and smoothed binarized maps of weights for 180000 pixels have been used. The pixels have been randomly selected from 100 different images taken randomly from different databases. A zero-mean Gaussian white noise was added to selected images with the following values of σ (5, 10 and 20).

Besides the neural network with 9 inputs, we have also considered the neural networks with smaller number of inputs. The results of experiments are presented in Table I.

TABLE I. RESULTS OF THE NEURAL NETWORK LEARNING FOR DIFFERENT SETS OF FEATURES

	Features					
	#1 and #4	#1, #2, #3	#4, #5, #6	#7, #8, #9	First 6	All 9
MSE of learning	0.154	0.197	0.149	0.198	**0.139**	**0.133**

As it is seen from Table I, there is a possibility to decrease a number of inputs to 6 with a slightly increased MSE of learning. However, this is not necessary, since the time for calculation of the neural network's output is insignificant in

ISBN 978-0-9928626-7-1

comparison to the time of calculation of feature maps #4, #5 and #6. Since these features are the most informative ones, their removal from inputs will significantly increase MSE of learning.

III. NUMERICAL ANALYSIS

Let us compare the performance of A-BM3D filter, its ability to provide lower filtered wMSE values (or bigger wPSNR values) than those by filters P-BM3D and S-BM3D.

For the analysis, three sets of test images are used. First, we have selected randomly 22 grayscale images in quarter resolution (384x256 pixels) from Amsterdam Library of Textures (ALOT) [9]. Second, grayscale versions of 25 reference images from TID2013 database [2] are use in the analysis. Finally, we have used the grayscale versions of 16 standard test images: Aerial, Airfield, Baboon, Barbara, Bikes, Bird, Boat, Bridge, Cameraman, Goldhill, Harbour, Lena, Livingroom, Man, Moon, Peppers.

For all these images a zero-mean Gaussian white noise with σ=15 was added. The value σ=15 (different from those values used for the neural network training) is selected to assure that the normalization of features maps is correct and that the proposed A-BM3D filter is able to deal with any value σ of the standard deviation of noise.

Fig. 4 shows curves of wPSNR for the filters A-BM3D, S-BM3D and P-BM3D for all three test sets. As it is can be seen from this Figure, there are images for which P-BM3D has larger wPSNR than S-BM3D, and for other images S-BM3D has larger wPSNR than P-BM3D. In all the cases A-BM3D provides largest value of wPSNR, sometimes outperforming S-BM3D and P-BM3D by about 1 dB.

Fig. 5 illustrates better visual quality of output of the proposed A-BM3D filter.

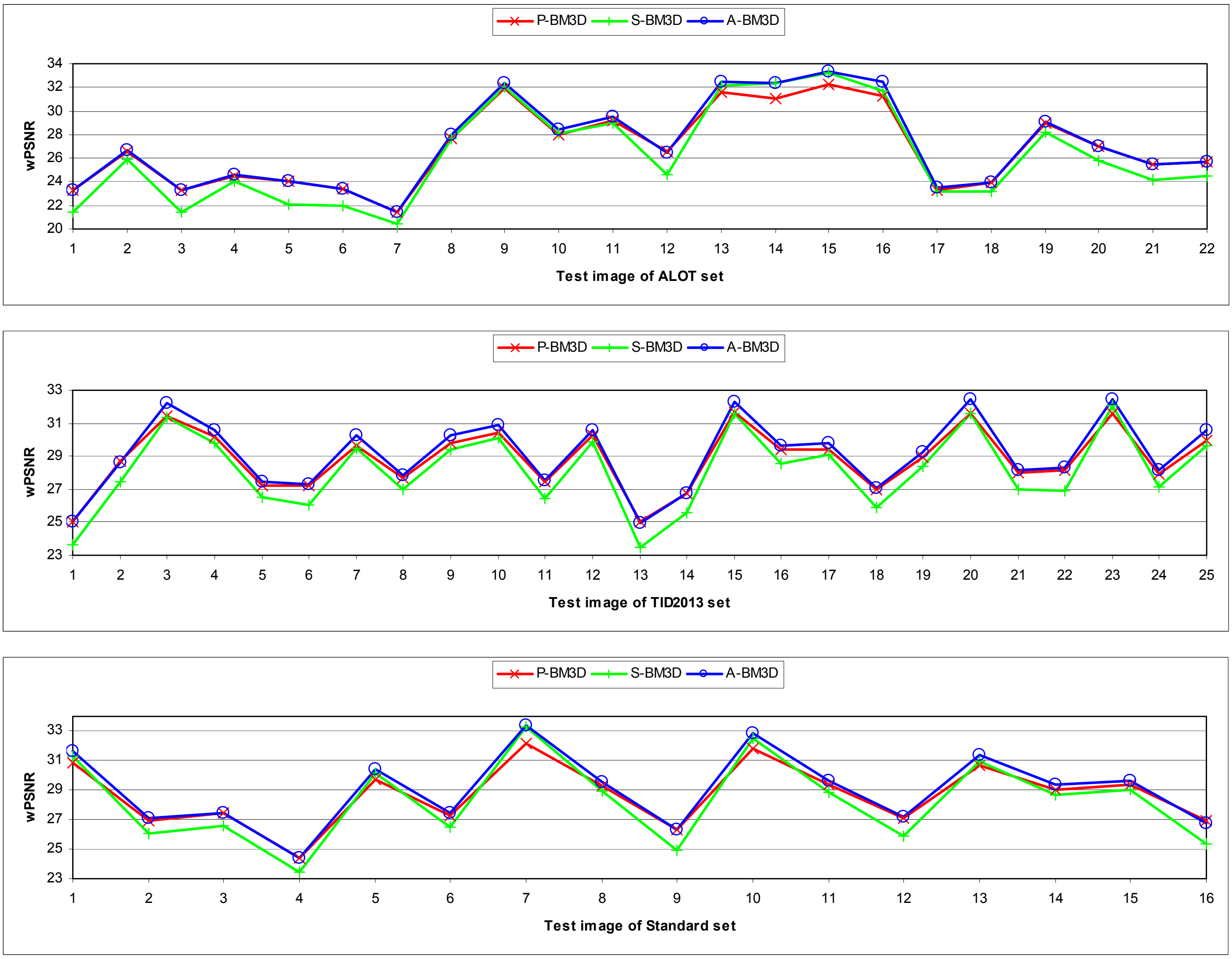
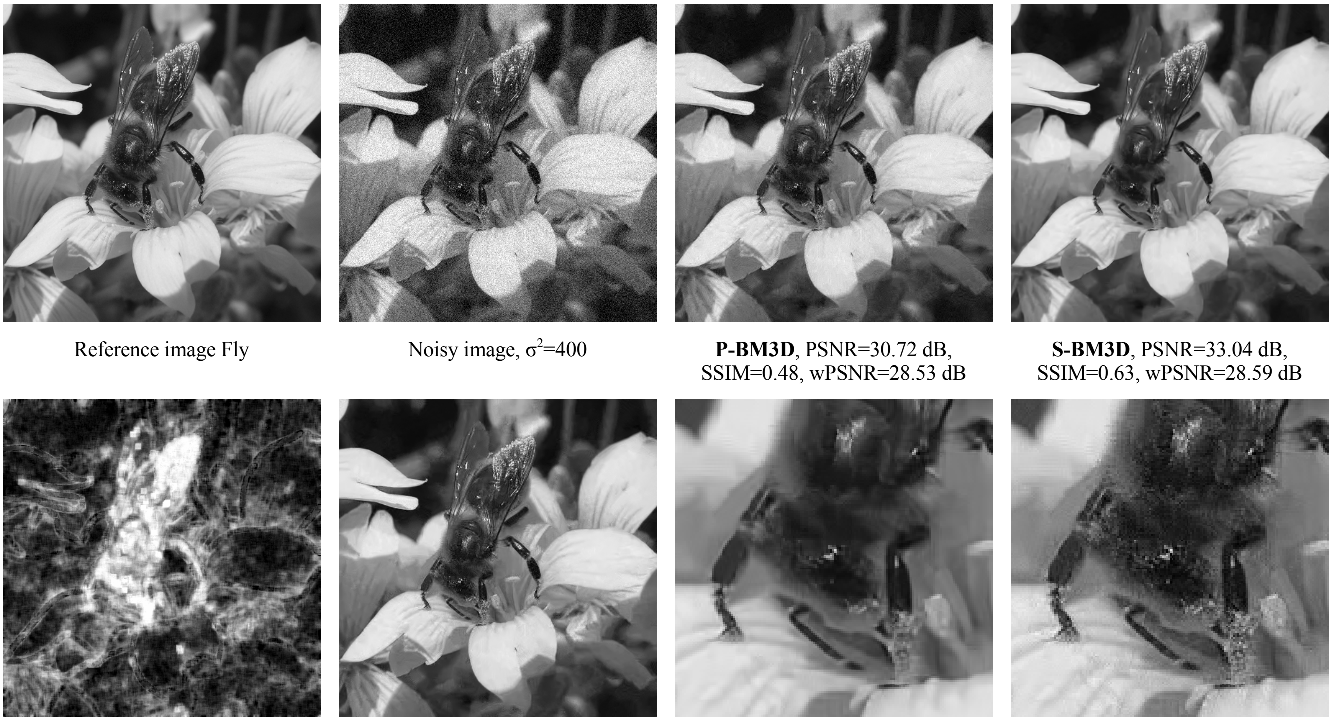


Fig. 4. Values of wPSNR for compared denoising methods for different sets of images (σ=15)

ISBN 978-0-9928626-7-1



Reference image Fly | Noisy image, σ^2=400 | **P-BM3D**, PSNR=30.72 dB, SSIM=0.48, wPSNR=28.53 dB | **S-BM3D**, PSNR=33.04 dB, SSIM=0.63, wPSNR=28.59 dB

Output of neural network | **A-BM3D**, PSNR=32.51 dB, SSIM=0.60, wPSNR=29.42 dB | Enlarged fragment of S-BM3D output | Enlarged fragment of A-BM3D output

Fig. 5. Visual comparison of the Fly test image denoising using P-BM3D, S-BM3D and proposed A-BM3D

As it can be seen, P-BM3D preserves better low contrast texture and fine details while S-BM3D suppresses better noise in homogeneous regions. At the same time, the proposed A-BM3D combines advantages of both BM3D profiles, providing spatially adaptivity between the profiles in the proper places of the image. It is also confirmed by the value of wPSNR metric which for A-BM3D output is larger than those for S-BM3D and P-BM3D ones by over 0.8 dB. Note that both widely used metrics PSNR and SSIM fail in this case, indicating wrongly that the oversmoothed output of S-BM3D provides superior quality than the output of A-BM3D.

Conclusions

In this paper, a new spatially adaptive modification of BM3D filter, A-BM3D was introduced. It alternates (switches) between two profiles of BM3D using specially designed neural network, output of which was calculated for each pixel of the filtered image. It has been demonstrated that A-BM3D provides lower values of wMSE, which was used to train the neural network.

Perspectives for a further study may consist in usage of more filter profiles in the design and in application of other image quality metrics with a better correspondence to a human perception than wMSE.

Acknowledgment

This work is supported by Academy of Finland, project no. 287150, 2015-2019.

References

[1] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing* 13(4), 2004, pp. 600–612.

[2] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives", *Signal Processing: Image Communication*, vol. 30, 2015, pp. 55-77.

[3] D. Chandler, "Seven Challenges in Image Quality Assessment: Past, Present and Future Research", ISNR *Signal Processing*, vol. 2913, 2013, pp. 1–53.

[4] N. Ponomarenko, S. Krivenko, K. Egiazarian, V. Lukin, J. Astola, "Weighted mean square error for estimation of visual quality of image denoising methods", CD ROM Proceedings of *VPQM*, 2010, 6 p.

[5] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, "New full-reference quality metrics based on HVS", Proceedings of the Second International Workshop on Video Processing and Quality Metrics, *VPQM*, 2006, 4p.

[6] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, "On between-coefficient contrast masking of DCT basis functions", Proceedings of the third international workshop on video processing and quality metrics, *VPQM*, 2007, 4 p.

[7] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering", *IEEE Transactions on Image Processing*, 16(8), 2007, pp. 2080-2095.

[8] S. Abramov, B. Vozel, J. Astola, K. Chehdi, V. Zabrodina, V. Lukin, "Methods for blind estimation of the variance of mixed noise and their performance analysis", *INTECH Open Access Publisher*, 2011.

[9] G. J. Burghouts and J. M. Geusebroek, Material-specific adaptation of color invariant features, *Pattern Recognition Letters*, vol. 30, 2009, pp. 306-313.