20th IMEKO TC4 International Symposium and
18th International Workshop on ADC Modelling and Testing
Research on Electric and Electronic Measurement for the Economic Upturn
Benevento, Italy, September 15-17, 2014

On the Accuracy of Vector Metrics for Quality Assessment in Image Filtering

Fabrizio Russo

Department of Engineering and Architecture – University of Trieste
Via A. Valerio 10, Trieste, Italy e-mail: rusfab@univ.trieste.it

***Abstract* – Noise cancellation and detail preservation are two conflicting goals in the design of any denoising filter. In order to specifically measure these features, *vector metrics* have been recently introduced for grayscale and color images. This paper provides, for the first time, a study of the accuracy of the main algorithms in the literature focusing on the case of Gaussian noise. A novel vector method is also proposed to overcome the inaccuracies of current techniques.**

Keywords: image processing, denoising filters, vector metrics, image quality.

I. INTRODUCTION

Vector metrics are a recently introduced class of full-reference metrics for performance evaluation of image denoising filters [1-4]. In this approach a vector error is computed whose components evaluate the different amounts of residual noise, color and structural distortion remaining after noise removal. Vector metrics overcome the limitations of classical operators, such as the mean squared error (MSE) and the peak signal-to-noise ratio (PSNR) that cannot discriminate between noise cancellation and detail preservation given by a filter. Vector methods also overcome the limitations of the perceptual quality metrics that can be insensitive to different combinations of unfiltered noise and image distortion [5-7].

In the last few years, different implementations of the *vector root mean squared error* (VRMSE) have been proposed in the literature. All of them split the MSE of the luminance channel of the image into two components that respectively deal with noise cancellation and detail preservation However, an in-depth analysis of the accuracy given by the various algorithms has not been presented yet.

The aim of this paper is twofold:

1) to provide a novel procedure for the study of the accuracy of this class of vector metrics;
2) to propose a new VRMSE method that is much more accurate than previous techniques.

This paper is organized as follows. Section II gives a brief overview of the available VRMSE techniques, Section III presents the new method, Section IV discusses the results of many computer simulations, and, finally, Section V reports the conclusions.

II. A BRIEF OVERVIEW OF VRMSE METHODS

Although the RGB color space is the most popular color coordinate system for image storage, display and processing, a different choice such as the YIQ color space [8-9] is more convenient if we want to separate the luminance error from the chrominance one. In the YIQ model, the luminance is represented by the Y channel, whereas the I and Q components deal with the chrominance information only. Thus, let $\mathbf{p}(\mathbf{n})=[p_Y(\mathbf{n}), p_I(\mathbf{n}), p_Q(\mathbf{n})]^T$, be the vector (in the YIQ space) representing the pixel at spatial position $\mathbf{n}=[n_1,n_2]$ in the original noise-free image and let $\mathbf{d}(\mathbf{n})=[d_Y(\mathbf{n}), d_I(\mathbf{n}), d_Q(\mathbf{n})]^T$ be the corresponding pixel in the filtered picture. Thus, the VRMSE can be defined as follows [3]:

$$\mathbf{VRMSE} = \left[\mathrm{RMSE}_{\mathrm{LUM}}, \mathrm{RMSE}_{\mathrm{CHR}}\right] \tag{1}$$

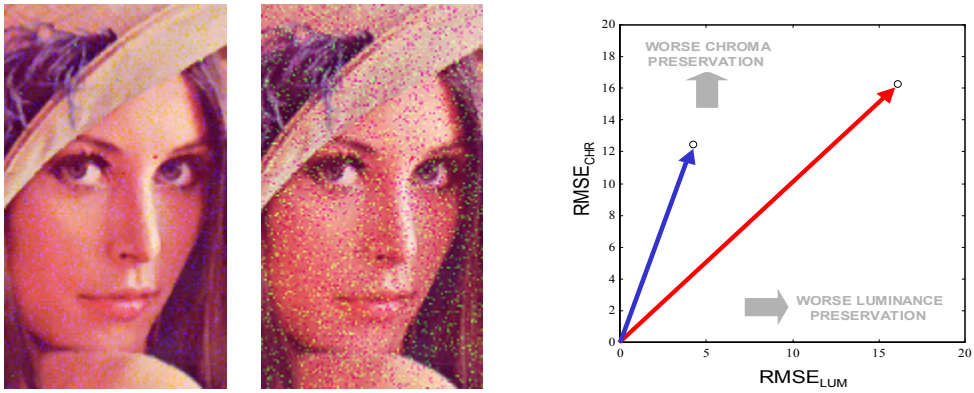
where the *luminance root mean squared error* $\mathrm{RMSE}_{\mathrm{LUM}}$ and the *chroma root mean squared error* $\mathrm{RMSE}_{\mathrm{CHR}}$ are given by the following relationships:

$$\mathrm{RMSE}_{\mathrm{LUM}} = \sqrt{\frac{1}{N}\sum_{\mathbf{n}}\left(d_Y(\mathbf{n}) - p_Y(\mathbf{n})\right)^2} \tag{2}$$

$$\mathrm{RMSE}_{\mathrm{CHR}} = \sqrt{\frac{1}{N}\sum_{\mathbf{n}}\left(d_I(\mathbf{n}) - p_I(\mathbf{n})\right)^2 + \left(d_Q(\mathbf{n}) - p_Q(\mathbf{n})\right)^2} \tag{3}$$

and N denotes the number of processed pixels. An example of the advantages of the YIQ choice is reported in Fig.1. Fig.1a and 1b show portions of filtered images with different visual quality but having almost the same MSE computed in the RGB space. Conversely, the quantitative evaluations in the YIQ space highlight the differences: $\mathrm{MSE}_{\mathrm{YIQ}}$=174.1 (Fig.1a), $\mathrm{MSE}_{\mathrm{YIQ}}$=525.2 (Fig.1b). The vector errors are depicted in Fig.1c.

ISBN-14: 978-92-990073-2-7



(a) (b) (c)

Fig.1. Portions of pictures having the same MSE_{RGB} but different MSE_{YIQ} evaluations: (a) MSE_{YIQ}=174.1, (b) MSE_{YIQ}=525.2, (c) graphical representation of the corresponding vector errors.

In general, the $RMSE_{LUM}$ can also be interpreted as a vector and then decomposed into two components, namely $RMSE_A$ and $RMSE_B$, that respectively deals with the noise cancellation and the detail preservation:

$$\mathbf{VRMSE_{LUM}} = [\mathrm{RMSE_A}, \mathrm{RMSE_B}] \tag{4}$$

It should be observed that the YIQ color space guarantees compatibility with monochrome pictures [8-9], so the $VRMSE_{LUM}$ decomposition defined by (4) can be directly applied to a grayscale image too.

Different implementations are available in the literature depending on the kind of noise and the algorithm that is adopted to compute the $RMSE_A$ and $RMSE_B$ components. The case of impulse noise is simple: such components can be evaluated in the set of pixel coordinates corresponding to corrupted and uncorrupted pixels, respectively. The case of Gaussian (or uniform) noise is more difficult because all image pixels are corrupted by noise. In general, a *type-j* $VRMSE_{LUM}$ decomposition is defined by the following pair of relationships:

$$\mathrm{RMSE_{Aj}} = \sqrt{\frac{1}{N}\sum_{\mathbf{n}} \chi_j(\mathbf{n})\left(d_Y(\mathbf{n}) - p_Y(\mathbf{n})\right)^2} \tag{5}$$

$$\mathrm{RMSE_{Bj}} = \sqrt{\frac{1}{N}\sum_{\mathbf{n}} \left(1-\chi_j(\mathbf{n})\right)\left(d_Y(\mathbf{n}) - p_Y(\mathbf{n})\right)^2} \tag{6}$$

where $\chi_j(\mathbf{n})$ is an appropriate function ($0 \le \chi_j \le 1$).

In type-1 VRMSE [3], the decomposition function $\chi_1(\mathbf{n})$ is obtained from a map of edge gradients of the original image and measures how much a given pixel belongs to an uniform region: $\chi_1(\mathbf{n}) = 1 - s_Y(\mathbf{n})$, where s_Y is the output of a classical 3×3 Sobel edge detector rescaled to the interval [0,1]. On the contrary, the type-2 VRMSE decomposition does not adopt any edge map.

It is based on the classification of filtering errors in two classes [4]; errors caused by insufficient filtering that does not completely remove the noise and errors caused by excessive (or wrong) filtering that produces unwanted distortion. According to (5)-(6), this operation is performed by a decomposition function that aims at separating pixels with residual noise from those affected by distortion:

$$\chi_2(\mathbf{n}) = \begin{cases} 1 & \text{if } p_Y(\mathbf{n}) < d_Y(\mathbf{n}) \le q_Y(\mathbf{n}) \\ 1 & \text{if } q_Y(\mathbf{n}) \le d_Y(\mathbf{n}) < p_Y(\mathbf{n}) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

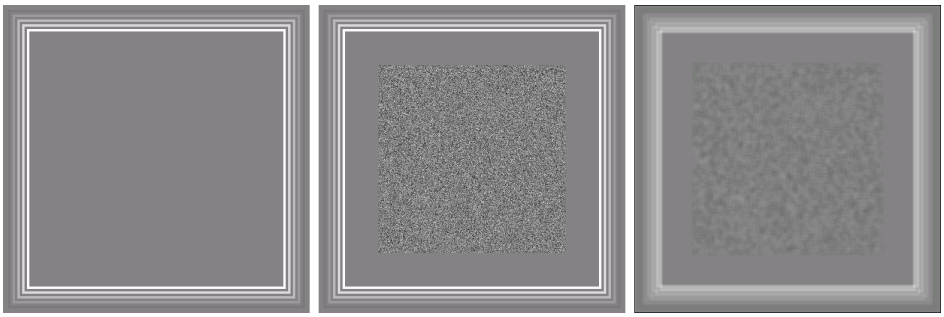
where $q(\mathbf{n})$ is the pixel luminance at location $\mathbf{n}=[n_1,n_2]$ in the noisy image.

III. INTRODUCING THE TYPE-3 $VRMSE_{LUM}$

As will be shown in the next Section, type-1 and type-2 methods are affected by some inaccuracies. A new method, called type-3 $VRMSE_{LUM}$, is here proposed to produce more accurate results. In this method, a novel decomposition function, namely $\chi_3(\mathbf{n})$, is defined as follows:

$$\chi_3(\mathbf{n}) = \begin{cases} 1 & \text{if } \Delta(\mathbf{n}) \le T \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where T is a threshold and $\Delta(\mathbf{n})$ represents the absolute difference between the pixel luminance $p_Y(\mathbf{n})$ at location $\mathbf{n}=[n_1,n_2]$ in the (noise-free) reference image and the pixel luminance $b_Y(\mathbf{n})$ that is obtained when this reference picture is filtered with the same parameter settings: $\Delta(\mathbf{n}) = |p_Y(\mathbf{n}) - b_Y(\mathbf{n})|$. Relation (8) performs a segmentation of the reference image into two crisp subsets of pixels: subset A_3 ($\chi_3=1$), where filtering distortion is not expected



(a) (b) (c)

Fig.2. Reference test image (a), noisy test image (b), result of 9×9 mean filtering.

and subset B_3, where such distortion does occur (χ_3=0). In the new method, the $RMSE_{A3}$ and $RMSE_{B3}$ components are evaluated in the mentioned subsets A_3 and B_3, respectively. Notice that, unlike other approaches, subset A_3 and B_3 depend on the actual filtering parameters (e.g. the window size), as it should be. An *offset-correction* procedure is adopted to improve the accuracy of the results and ensure that a single choice of T yields satisfactory results for reference pictures that are not perfectly noiseless (as commonly occurs when real image data are considered). Let MSE_{A30} be the possibly nonzero offset component that is achieved when the reference image is filtered. Thus, the components $RMSE_{A3}=(MSE_{A3})^{1/2}$ and $RMSE_{B3}=(MSE_{B3})^{1/2}$ are updated as described by the following C pseudocode:

```
if (MSE_A30<MSE_A3) {
        MSE_B3=MSE_B3+MSE_A30;
        MSE_A3=MSE_A3−MSE_A30;
}
else {
        MSE_B3=MSE_B3+MSE_A3;
        MSE_A3=0;
}
```

The results of a heuristic search dealing with many test images show that T=15 is a satisfactory choice for a wide range of noise variances.

IV. PERFORMANCE EVALUATION OF $VRMSE_{LUM}$ METHODS

Since all $VRMSE_{LUM}$ methods process pixel luminances, i.e., grayscale values, we shall consider grayscale test images in the following experiments.

For the first group of tests, we generated the noise-free 512×512 reference image shown in Fig.2a. In this picture, all edges are located close to the borders, whereas the rest of the image is perfectly uniform. If we add noise to the central region only, we shall know the true values of $RMSE_A$ and $RMSE_B$. Thus, we produced a noisy picture by adding zero-mean Gaussian noise with standard deviation σ=40 (Fig.2b), and we adopted mean filters with different window sizes to increase the distortion (the result given by 9×9 mean filtering is depicted in Fig.2c). The lists of $RMSE_A$ values given by type-1 ($RMSE_{A1}$) [3], type-2 ($RMSE_{A2}$) [4] and type-3 ($RMSE_{A3}$) metrics are reported in Table 1. The $RMSE_B$ values yielded by type-1 ($RMSE_{B1}$) [3], type-2 ($RMSE_{B2}$) [4] and type-3 ($RMSE_{B3}$) methods are listed in Table 2. Observing the data in Table 1, we can see that only the novel method (fifth column) gives the exact values (second column), whereas other techniques yield less accurate evaluations. The results in Table 2 lead to the same considerations. In the second group of experiments we considered three 512×512 grayscale images: "Lena", "Lighthouse" and "House" (Fig.3). We generated three noisy pictures by adding zero-mean Gaussian noise with standard deviation σ=15.

Table 1. List of $RMSE_A$ evaluations.

Mean Filter	$RMSE_A$ true	$RMSE_{A1}$ [3]	$RMSE_{A2}$ [4]	$RMSE_{A3}$ proposed
5-point	10.90	10.95	7.77	10.90
3×3	8.12	8.28	5.76	8.12
5×5	4.88	7.48	3.45	4.88
7×7	3.48	11.63	2.46	3.48
9×9	2.70	14.75	1.91	2.70

Table 2. List of $RMSE_B$ evaluations.

Mean Filter	$RMSE_B$ true	$RMSE_{B1}$ [3]	$RMSE_{B2}$ [4]	$RMSE_{B3}$ proposed
5-point	5.32	5.23	9.32	5.32
3×3	8.80	8.65	10.50	8.80
5×5	11.85	10.40	12.34	11.85
7×7	15.76	11.20	15.95	15.76
9×9	20.24	14.12	20.33	20.24

(a) (b) (c)

Fig.3. Test images: (a) Lena, (b) Lighthouse, (c) House.

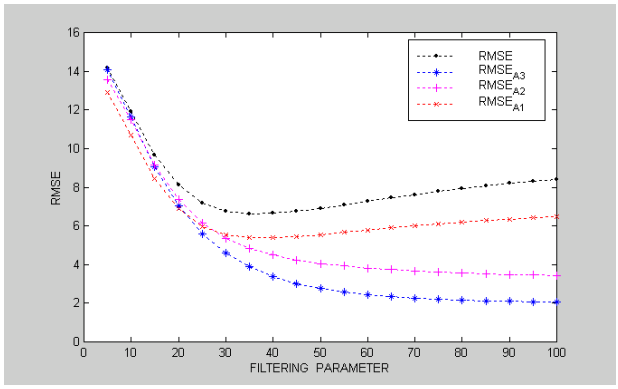
We chose a 7×7 bilateral filter because its behavior is well known [10-11]: the noise cancellation and the image distortion increase as the value of the main parameter σ_r increases. The results of bilateral filtering when the parameter σ_d is assigned (σ_d=5) and σ_r ranges from 5 to 100 are shown in Figs.4, 5, and 6. The $RMSE_{LUM}$ values are also reported for a comparison (black lines, briefly RMSE). Let us focus on the data in Fig.4a: as the value of the filtering parameter σ_r increases, the remaining noise decreases and the same should occur for the $RMSE_A$ component that evaluates this effect. The novel $RMSE_{A3}$ (blue line) operates according to this behavior: it decreases rapidly toward very low values that are correctly obtained when almost all the noise has been filtered. The $RMSE_{A2}$ (magenta line) decreases less quickly. The $RMSE_{A1}$ (red line) shows an incorrect behavior: it increases for σ_r>40. It should be observed that an incorrect $RMSE_A$ could be caused by a wrong evaluation of the corresponding $RMSE_B$ component, because $(RMSE_{LUM})^2=(RMSE_A)^2+(RMSE_B)^2$. In this respect, the graphical representation of the $RMSE_B$ values is very useful (Fig.4b). Clearly, the $RMSE_B$ should increase as the filtering parameter σ_r (and so the unwanted distortion) increases. Only the proposed $RMSE_{B3}$ performs according to this behavior. It increases and becomes very close to the overall RMSE for large values of σ_r, when the distortion is almost the only effect of the filtering. On the contrary, the $RMSE_{B2}$ decreases for σ_r<10, and the $RMSE_{B1}$ behaves even worse (it reaches a minimum for about σ_r=35). The data represented in Fig.5 and 6 lead to the same conclusions and similarly does a third group of tests dealing with σ=30 (Fig.7 and 8). Notice that the mentioned choice T=15 was adopted for all the experiments. The erroneous behavior of type-1 VRMSE can be explained as follows. When strong filtering is applied, some relevant blur can cross the borders of the edge regions thus affecting the adjacent uniform areas where the $RMSE_{A1}$ is computed. In this case, $RMSE_{A1}$ and $RMSE_{B1}$ are respectively over- and under-estimated. As an example, the fraction of blur erroneously transferred to $RMSE_{A1}$ (Fig.4a) becomes dominant for σ_r>40 and then the overall $RMSE_{A1}$ increases instead of decreasing. Conversely, when weak filtering is performed, a significant fraction of unfiltered noise remains on the image edges where the $RMSE_{B1}$ is evaluated. In this case, $RMSE_{A1}$ and $RMSE_{B1}$ are under- and over-estimated, respectively. Weak filtering also highlights the inaccuracy (over-estimate) of the $RMSE_{B2}$ component in type-2 VRMSE. Indeed, filtering errors occurring in uniform areas can wrongly be classified as distortion by (7). Type-3 VRMSE overcomes all the limitations of the previous vector methods because the novel decomposition function $\chi_3(\mathbf{n})$ determines an adaptive behavior. It basically operates as follows**.** When weak filtering is applied, the set B_3 includes a few pixels. As a result, $RMSE_{B3}$ is not over-estimated and $RMSE_{A3}$ correctly encompasses most filtering errors produced by noise smoothing. When strong filtering is applied, the set B_3 increases. The $RMSE_{B3}$ is not under-estimated because it can include all the pixels affected by the filtering distortion.

V. CONCLUSIONS

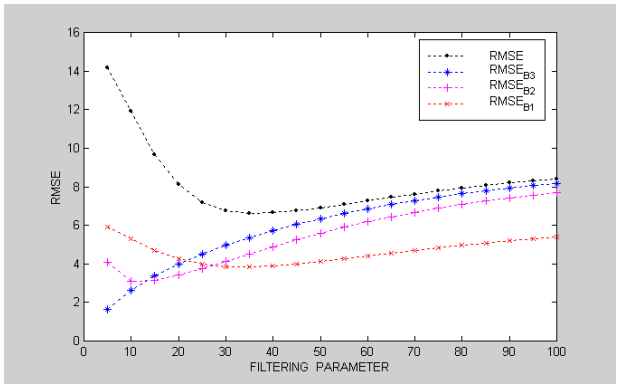
A novel procedure to evaluate the performance of vector metrics has been reported in this paper. The study has revealed that current vector techniques are not very accurate when they estimate the residual noise and the collateral distortion produced by a denoising filter. To address this issue, a novel vector method, called *type-3 VRMSE*, has been presented. The evaluation procedure has shown that the new method exhibits a satisfactory behavior and performs much better than other existing vector techniques.

REFERENCES

[1] A. De Angelis, A. Moschitta, F. Russo, P. Carbone, "Image Quality Assessment: An Overview and Some Metrological Considerations", in Proc. IEEE Workshop AMUEM, Sardagna, Italy, Jul. 16-18, 2007, pp. 47-52.

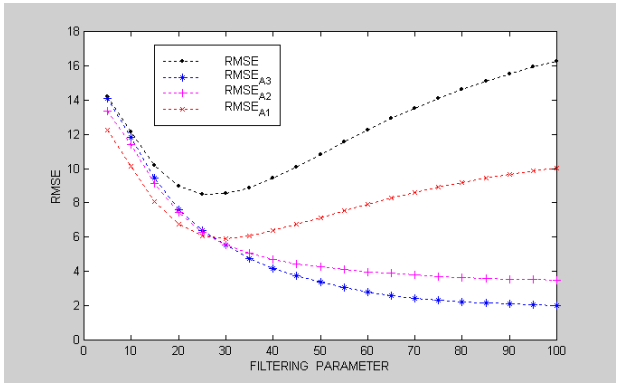


(a)

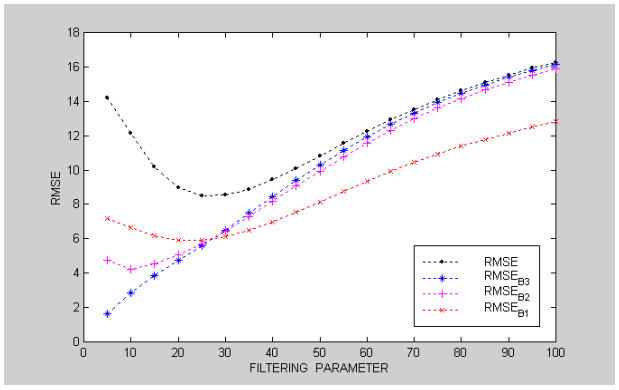


(b)

Fig. 4. $RMSE_A$ (a) and $RMSE_B$ (b) evaluations for the "Lena" image corrupted by Gaussian noise (σ=15) and processed by bilateral filtering (σ_d=5, $5 \le \sigma_r \le 100$).

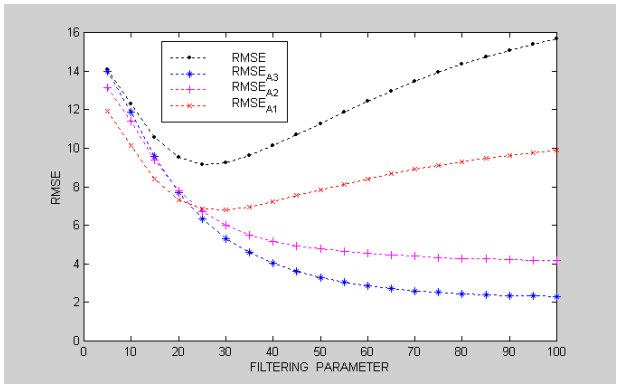


(a)

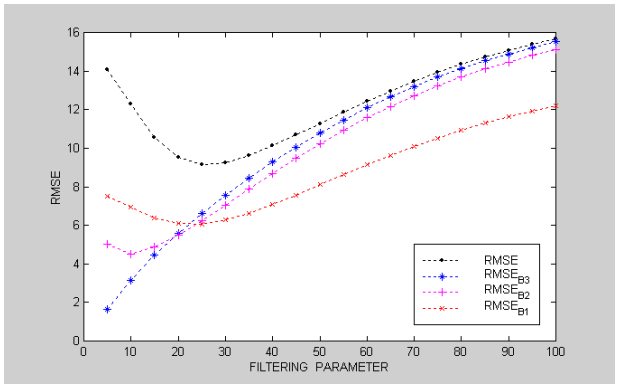


(b)

Fig.5. $RMSE_A$ (a) and $RMSE_B$ (b) evaluations for the "Lighthouse" image corrupted by Gaussian noise (σ=15) and processed by bilateral filtering (σ_d=5, $5 \le \sigma_r \le 100$).

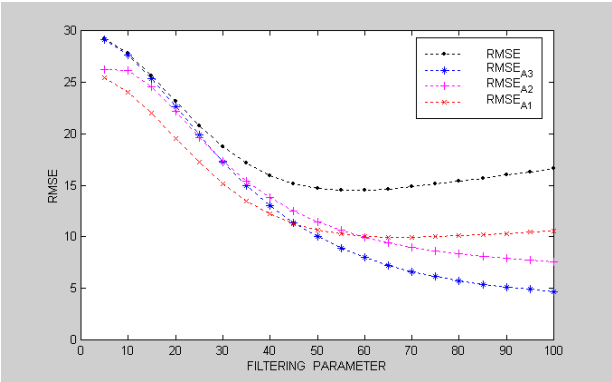


(a)

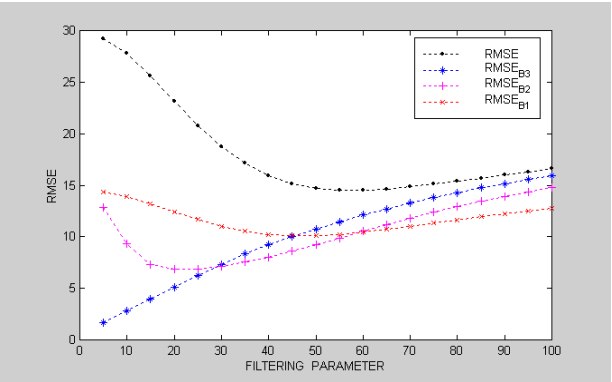


(b)

Fig.6. $RMSE_A$ (a) and $RMSE_B$ (b) evaluations for the "House" image corrupted by Gaussian noise (σ=15) and processed by bilateral filtering (σ_d=5, $5 \le \sigma_r \le 100$).

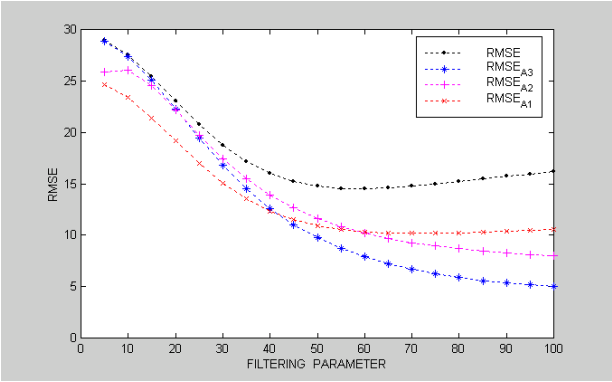


(a)

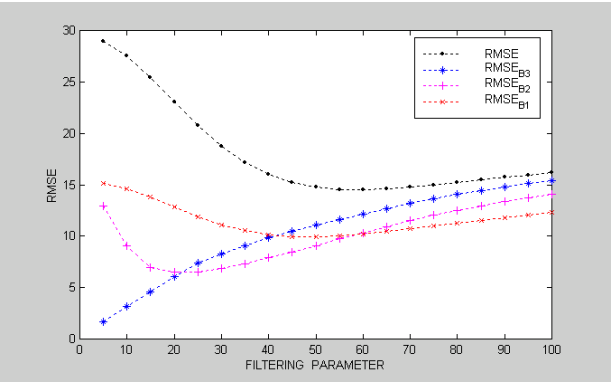


(b)

Fig. 7. $RMSE_A$ (a) and $RMSE_B$ (b) evaluations for the "Lighthouse" image corrupted by Gaussian noise (σ=30) and processed by bilateral filtering (σ_d=5, $5 \le \sigma_r \le 100$).



(a)



(b)

Fig. 8. $RMSE_A$ (a) and $RMSE_B$ (b) evaluations for the "House" image corrupted by Gaussian noise (σ=30) and processed by bilateral filtering (σ_d=5, $5 \le \sigma_r \le 100$).

[2] F. Russo, A. De Angelis, P. Carbone, "A Vector Approach to Quality Assessment of Color Images", in Proc. I2MTC-2008, Victoria, Vancouver Island, Canada, May 12-15, 2008.

[3] A. De Angelis, A. Moschitta, F.Russo, P. Carbone, "A Vector Approach for Image Quality Assessment and Some Metrological Considerations", IEEE Trans. on Instrum. and Meas., vol.58, n.1, January 2009, pp.14-25.

[4] F. Russo, "New Method for Performance Evaluation of Grayscale Image Denoising Filters", IEEE Signal Process. Lett., vol.17, n.5, May 2010, pp.417-420.

[5] Z. Wang and A. C. Bovik, "A universal image quality index" IEEE Signal Process. Lett., vol. 9, no. 3, pp. 81-84, March 2002.

[6] D.-O. Kim and R.-H. Park, "New image quality metric using the Harris response", IEEE Signal Process. Lett., vol. 16, no. 7, pp. 616-619, 2009.

[7] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, V. Lukin "Metrics performance comparison for color image database", Fourth international workshop on video processing and quality metrics for consumer electronics, Scottsdale, Arizona, USA. Jan. 14-16, 2009.

[8] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", Pearson, 2008.

[9] W. Burger and M.J. Burge, "Digital Image Processing: An Algorithmic Introduction Using Java", Springer, 2012.

[10] M. Zhang and B. K. Gunturk, "Multiresolution Bilateral Filtering for Image Denoising", IEEE Transactions on Image Processing, vol.17, n.12, December 2008, pp.2324-2333.

[11] B. K. Gunturk, "Bilateral Filter: Theory and Application" in "Computational Photography: Methods and Applications", Lukac R. (Ed.), CRC Press, 2011, pp.339-366.