

Assessment of the spatio-temporal vector median filtering algorithms using colour video quality metrics

Krzysztof Okarma

West Pomerania University of Technology

71-126 Szczecin, ul. 26. Kwietnia 10, e-mail: okarma@zut.edu.pl

In this paper the application of the combined video quality assessment method as well as some other recently developed objective metrics for the analysis of the results of the nonlinear colour video filtering is discussed. The spatio-temporal versions of colour image filtering methods, including the Vector Median Filter, can be obtained using frame-by-frame approach but the proper choice of the spatio-temporal kernel weights and the colour space used during filtration should be based on a reliable video quality assessment. In some earlier papers the combined video quality assessment method has been proposed, which has a highly linear correlation with subjective quality scores and can be extended into the colour version. As the illustration of the problem, some results of the colour video denoising using the spatio-temporal VMF, also in a weighted version, together with the quality assessment results have been presented in the paper.

1. Introduction

Colour image processing is one of the most dynamically developing area of image processing and analysis. One of the most important reasons is the hardware progress, not only related to the colour image acquisition equipment, but also specialized graphical processing units allowing fast parallel processing of large amount of image data. Such progress requires some advances in colour image processing algorithms, which are not always a straightforward extension of well-known techniques used for greyscale imaging. A good example is nonlinear filtration of colour images, in particular the vector approach to median filtration.

Another relevant area of research is reliable quality assessment of colour images as well as colour video sequences. Most of quality metrics, even those recently proposed, do not use colour data at all. Since colour dissimilarity is an important element of image quality, the minimal requirement should be the calculation of an image metric for three RGB channels or additionally for the chrominance, depending on the colour model.

2. Nonlinear filtering of colour images

Classical approach to greyscale median filtering is used for the elimination of an impulse noise. Analysing the local neighbourhood of the current pixel, e.g. using the 3×3 pixels sliding window (mask), the most typical pixel from the mask

should be chosen as the output. For this purpose all the pixels within the current mask are sorted and the central (median) element is chosen. Such approach is characterised by an important advantage - the resulting image contains only the pixels' values which have been present in the original one.

In order to prevent such feature various versions of colour pixel's ordering can be used. A common idea is the choice of the most typical pixel from those which are present within the current mask. For this reason no additional colours are present in the resulting image.

One of possible approaches is the usage of the conditional ordering [3, 10], where the pixels are sorted according to a single, arbitrary chosen, colour channel. Only the pixels having the same value of that channel are additionally sorted according to the second one. Such approach can be used e.g. using the HSV colour model, using the value and saturation as first two ordering criterions, respectively.

Another group of colour image filtering methods is reduced (aggregate) ordering, where the calculation of the additional values based on all channels is necessary. Such values are used for ordering the pixels and the most typical one is chosen as the output.

Another well-known algorithm for nonlinear filtration of colour images is Vector Median Filter (VMF) [1], where the aggregated distance from each of the pixels to the remaining ones within the current mask is calculated in the specified colour space. As the result the pixel with the minimum aggregated distance (usually calculated using the L2-norm) is chosen as the most similar to the other ones. Instead of using the RGB colour space some other ones such as CIE XYZ or CIE L*a*b* can also be used [2].

3. Spatio-temporal video filtering

Taking into consideration a specific character of the video data a relatively large amount of noise can be observed, which is present for specified pixels in single video frames. Such noise can be removed using independent filtering of each frame, e.g. using Vector Median Filter, but for a large amount of an impulse noise the obtained results are often unsatisfactory. A typical approach used for the noise reduction from image sequences is the application of simple low-pass Arithmetic Mean Filter (AMF), which unfortunately causes easily noticeable blurring and loss of sharpness. Some other recently proposed ideas are based on sophisticated spatio-temporal filters based on motion estimation, which can be relatively easily applied for MPEG compressed sequences.

Nevertheless, the application of any spatio-temporal filter requires the proper choice of a spatio-temporal kernel (mask) containing the analysed pixels and their weights (used for the weighted filters). The choice of the kernel size and weights should be performed in a way leading to the maximum possible quality of the resulting image. Unfortunately, the optimisation is often performed using

traditional MSE-based image quality metrics, which are poorly correlated with subjective evaluations. For this reason some modern image quality assessment methods, such as Structural Similarity (SSIM) [11] should be used. Taking into account the fact, that the kernel size and weights can be defined for each RGB channel (or even using another colour space) separately, the colour image quality index should be used (or in the last resort three quality metrics' values for each channel).

Recently some other modern approaches to objective image and video quality assessment have been reported. Probably the most interesting ones are Multi-Scale SSIM [12], Visual Information Fidelity (VIF) [9] and R-SVD metric [4] based on the Singular Value Decomposition.

The MS-SSIM metric can be computed as the weighted product of three components for M (typically five) scales according to the following formula:

$$MS-SSIM(x, y) = [l(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c(x, y)]^{\beta_j} \cdot [s(x, y)]^{\gamma_j}, \quad (1)$$

where the luminance (l), contrast (c) and structural (s) distortions are defined similarly as for the single-scale SSIM:

$$SSIM(x, y) = \left(\frac{2\bar{x}\bar{y} + C_1}{\bar{x}^2 + \bar{y}^2 + C_1} \right)^{\alpha} \cdot \left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right)^{\beta} \cdot \left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \right)^{\gamma} = [l(x, y)]^{\alpha} \cdot [c(x, y)]^{\beta} \cdot [s(x, y)]^{\gamma}. \quad (2)$$

For both metrics the weighting coefficients (α, β, γ) are typically set to 1 for simplicity.

The VIF metric is defined as:

$$VIF = \frac{\sum_{j=0}^S \sum_{i=1}^{N_j} I(c_{i,j}; f_{i,j})}{\sum_{j=0}^S \sum_{i=1}^{N_j} I(c_{i,j}; e_{i,j})}. \quad (3)$$

where $I(a; b)$ denotes the mutual information between a and b . The denominator and numerator denote the information extracted by human vision from the reference and distorted images. Assuming that c is a block vector at a given location in the reference image, e denotes the perception of block c by a human viewer in the presence of an additive noise and f is the perception of distorted block c [9].

Another interesting approach to image quality assessment is the usage of the Singular Value Decomposition e.g. utilising the recently proposed R-SVD index calculated as:

$$R-SVD = \sqrt{\frac{\sum_{i=1}^{N_k} (d_i - 1)^2}{\sum_{i=1}^{N_k} (d_i + 1)^2}}, \quad (4)$$

where d_i denotes the singular values of the referee matrix $R^2 = U^2 \cdot \Lambda \cdot V^T$, with identity matrix Λ . The product of $U \cdot S \cdot V^T$ is obtained as the result of the SVD decomposition of the matrix representing the original image A , and $U^2 \cdot S^2 \cdot V^T$ is obtained for the distorted image A' respectively [4].

In one of the author's earlier papers [5] a combined image quality metric utilizing those metrics has been proposed and verified experimentally as highly linear correlated with subjective scores. Such metric can be expressed as:

$$CQM = (MS - SSIM)^a \cdot (VIF)^b \cdot (R - SVD)^c \quad (5)$$

assuming the nearly optimal simple combination of $a = 7$, $b = 0.3$ and $c = -0.15$ as the result of the optimisation in terms of the linear correlation with subjective quality scores using the TID2008 database [8], being the largest publicly available image database containing the subjective quality scores for 1700 colour images with many types of distortions.

Such approach has been further extended for the video quality assessment [6], as well as for the colour image quality assessment [7]. Nevertheless, in this paper an independent calculation of the metric for the RGB channels is used for convenient comparison with results obtained for the other quality metrics.

Some other interesting approaches, presented recently, are the application of the Riesz transform leading to the RFSIM metric [13], as well as the phase congruency and gradient maps resulting in the Feature Similarity (FSIM) metric [14]. Both of those metrics have good properties in the aspect of correlation with subjective evaluations, so they have also been used in the paper.

The RFSIM metric is defined as:

$$RFSIM = \prod_{i=1}^5 \frac{\sum_x \sum_y d_i(x, y) \cdot M(x, y)}{\sum_x \sum_y M(x, y)}, \quad (6)$$

where M stands for the binary mask obtained as the result of the morphological dilation of the image being the result of the Canny filter applied to the input image. The mask is multiplied pixel by pixel with the feature similarity calculated for each pixel with (x, y) coordinates as:

$$d_i(x, y) = \frac{2 \cdot f_i(x, y) \cdot g_i(x, y) + c}{f_i^2(x, y) + g_i^2(x, y) + c}, \quad i = 1..5, \quad (7)$$

where f_i and g_i are the five consecutive coefficients of the Riesz transform of the first and second order for the reference and distorted images.

The FSIM metric utilises the gradient map and the phase congruency instead of Riesz transform, so the overall metric can be expressed as:

$$FSIM = \frac{\sum_x \sum_y S(x,y) \cdot PC_m(x,y)}{\sum_x \sum_y PC_m(x,y)}, \quad (8)$$

where $PC_m(x,y) = \max(PC_1(x,y), PC_2(x,y))$, so the higher from the two values of the phase congruency for the reference and distorted image is used instead of the mask value. The local similarity S can be determined as:

$$S(x,y) = \left(\frac{2 \cdot PC_1(x,y) \cdot PC_2(x,y) + T_{PC}}{PC_1^2(x,y) + PC_2^2(x,y) + T_{PC}} \right)^\alpha \cdot \left(\frac{2 \cdot G_1(x,y) \cdot G_2(x,y) + T_G}{G_1^2(x,y) + G_2^2(x,y) + T_G} \right)^\beta \quad (9)$$

assuming the presence of two small stability constants T_{PC} and T_G and $\alpha = \beta = 1$ for simplicity.

The filtering results obtained in the experiments have been assessed using most state-of-the-art full-reference image quality assessment methods mentioned above, including the combined metric, as well as the PSNR for comparison purposes.



Fig. 1. Example video frames with their zoomed fragments (a – original, b – 30% impulse noise after VMF, c – 30% impulse noise after unweighted spatio-temporal VMF)

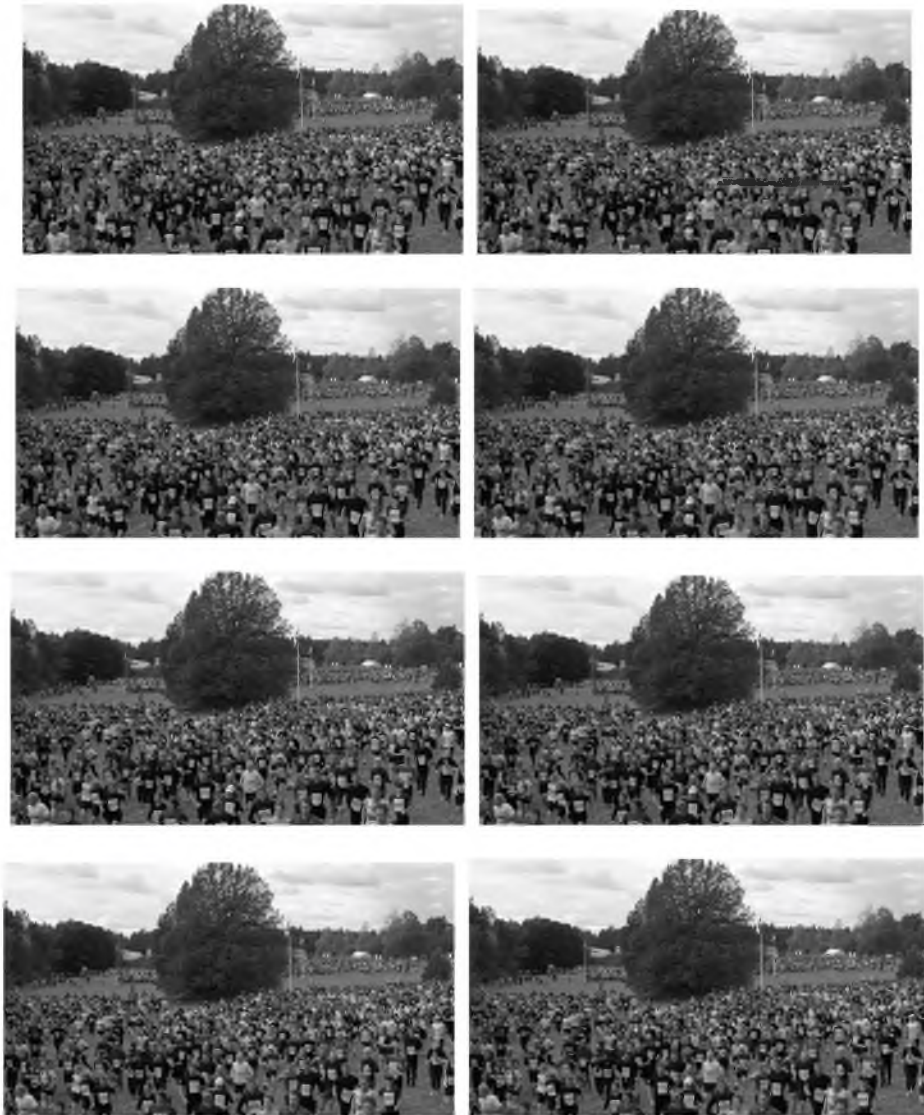


Fig. 2. Eight example consecutive frames of the reference test video

4. Experimental results

In order to illustrate the problem, some experiments have been performed using the 27-pixels spatio-temporal Vector Median Filter (3×3 pixels mask for 3 video frames) for the elimination of an RGB impulse noise from the well-known test sequence “Crowd run” characterised by many changes in a lower part of each frame. Exemplary video frames from the video sequence “Crowd run” used during

experiments are presented (luminance channel only) in Fig. 2. and some of the results obtained for one of the noised and filtered frames are presented in Fig. 1 with zoomed critical fragments of images. After the filtration, the image quality metrics' values for each RGB channel have been calculated and compared to those obtained using standard single-frame Vector Median Filter. The results presented as the average quality metrics' values for eight consecutive frames are presented in Figs. 3-10.

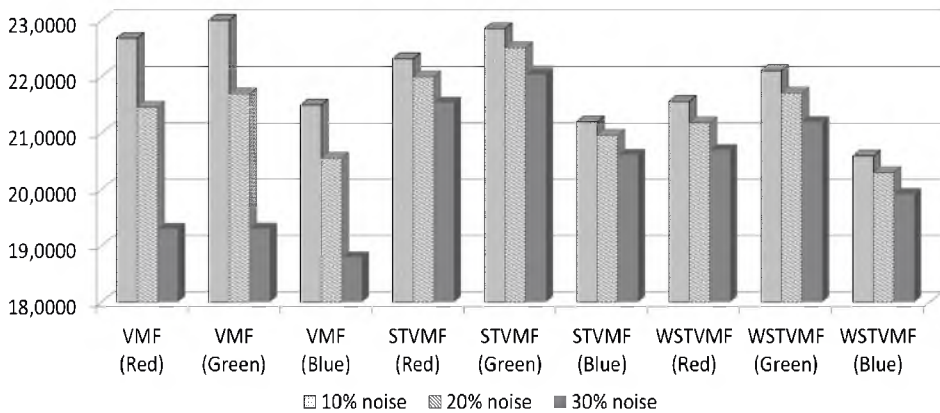


Fig. 3. PSNR values for the Vector Median Filter (VMF), spatio-temporal Vector Median Filter (STVMF) and weighted spatio-temporal one (STVMF) averaged for all frames

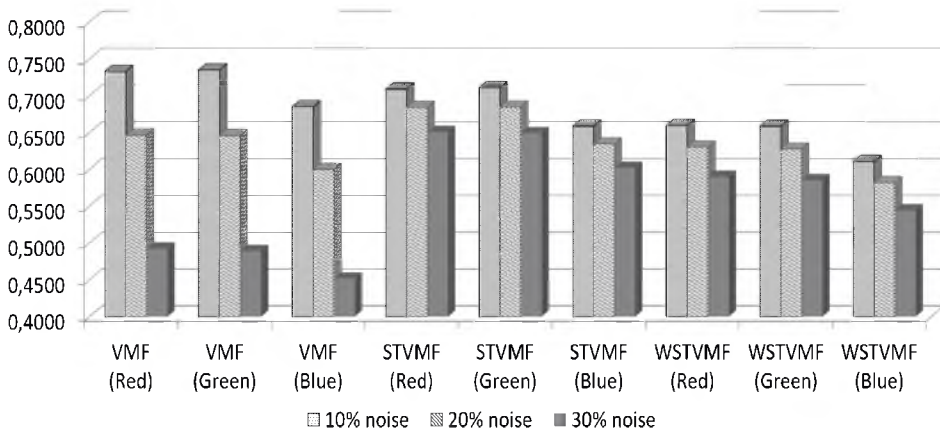


Fig. 4. Structural Similarity (SSIM) values for the Vector Median Filter (VMF), spatio-temporal Vector Median Filter (STVMF) and weighted spatio-temporal one (STVMF) averaged for all frames

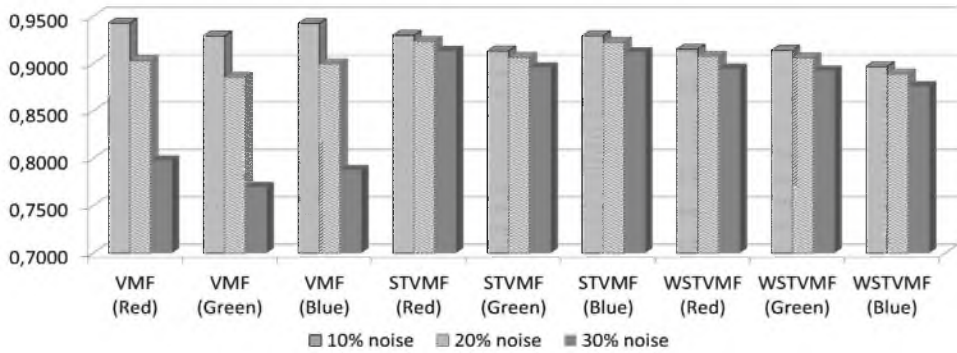


Fig. 5. Multi-Scale Structural Similarity (MS-SSIM) values for the Vector Median Filter (VMF), spatio-temporal Vector Median Filter (STVMF) and weighted spatio-temporal one (STVMF) averaged for all frames

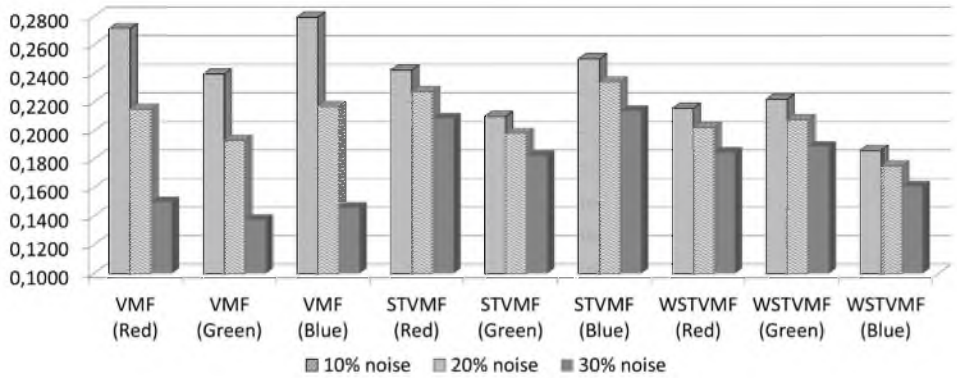


Fig. 6. VIF values for the Vector Median Filter (VMF), spatio-temporal Vector Median Filter (STVMF) and weighted spatio-temporal one (STVMF) averaged for all frames

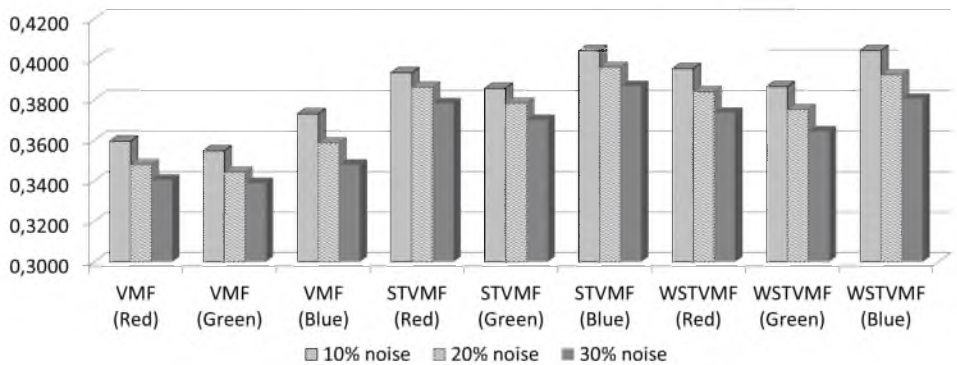


Fig. 7. R-SVD values for the Vector Median Filter (VMF), spatio-temporal Vector Median Filter (STVMF) and weighted spatio-temporal one (STVMF) averaged for all frames (smaller values indicate better quality)

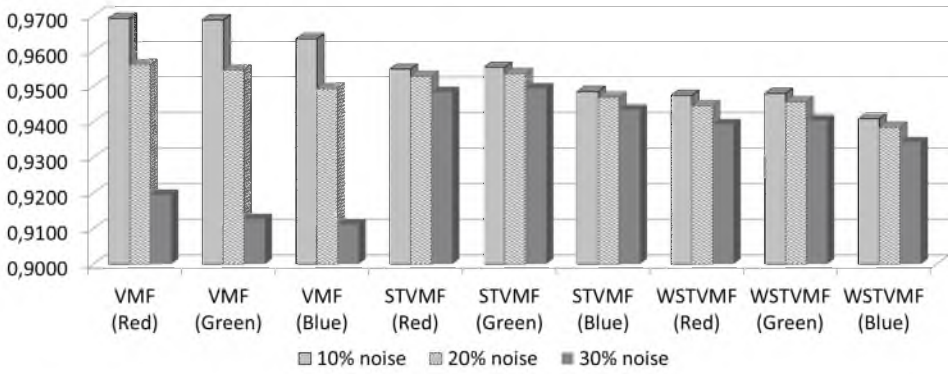


Fig. 8. FSIM values for the Vector Median Filter (VMF), spatio-temporal Vector Median Filter (STVMF) and weighted spatio-temporal one (STVMF) averaged for all frames

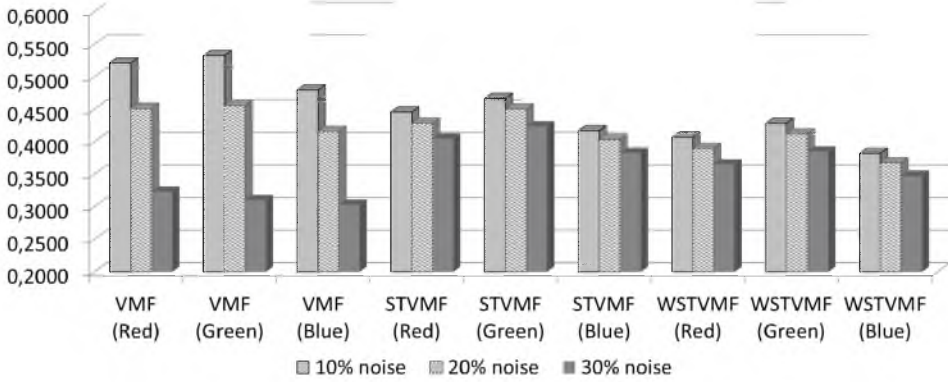


Fig. 9. RFSIM values for the Vector Median Filter (VMF), spatio-temporal Vector Median Filter (STVMF) and weighted spatio-temporal one (STVMF) averaged for all frames

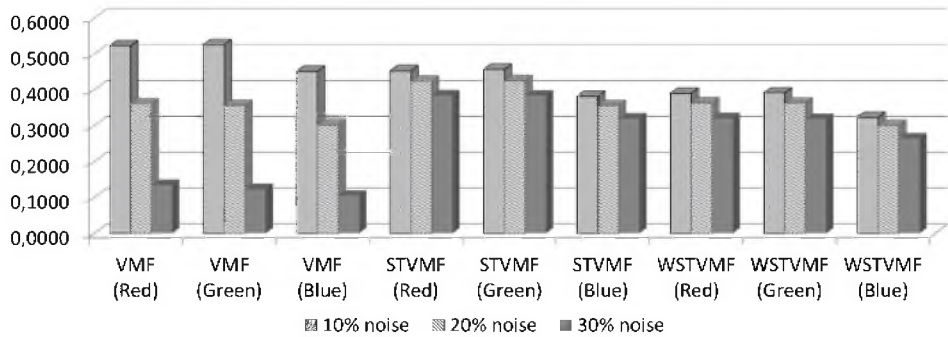


Fig. 10. Combined Quality Metric (CQM) values for the Vector Median Filter (VMF), spatio-temporal Vector Median Filter (STVMF) and weighted spatio-temporal one (STVMF) averaged for all frames

Analysing presented results the advantages of using the spatio-temporal VMF can be noticed, also for the combined metric, especially for the static or slow-motion fragments of the video frames with large amount of noise, even using the unweighted kernel. The improper choice of the spatio-temporal kernel weights may lead to worse results than the application of unweighted spatio-temporal filter. Most of the image quality metrics used in the paper lead to the correct conclusions using the frame-by-frame analysis, except from the R-SVD metric.

A more detailed analysis can be conducted calculating the values of the chosen quality metrics for the upper and lower parts of the frames separately. Since the upper part is rather static and the upper one is much more dynamic (large number of the nonzero motion vectors), the results obtained for those fragments of the images differ significantly. In order to limit the amount of presented data, only the MS-SSIM, FSIM and combined metric have been chosen for further analysis.

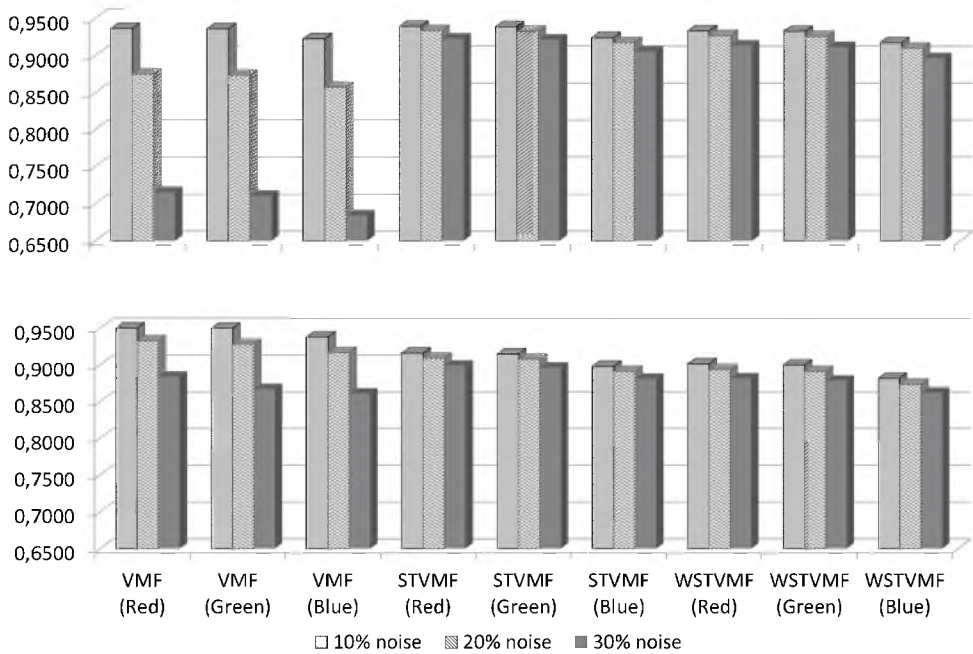


Fig. 11. MS-SSIM values for the upper and lower parts of the frames

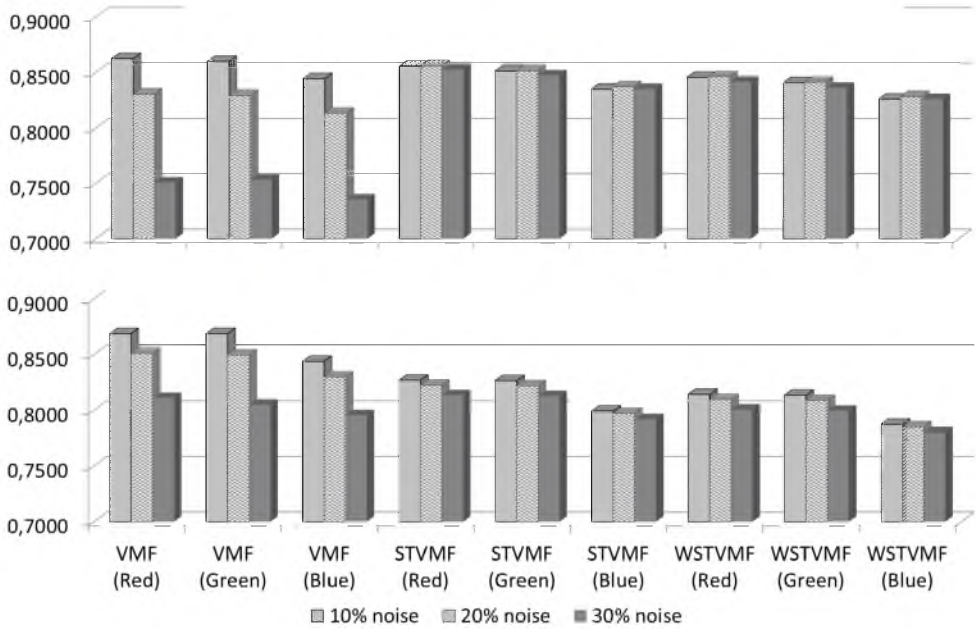


Fig. 12. FSIM values for the upper and lower parts of the frames

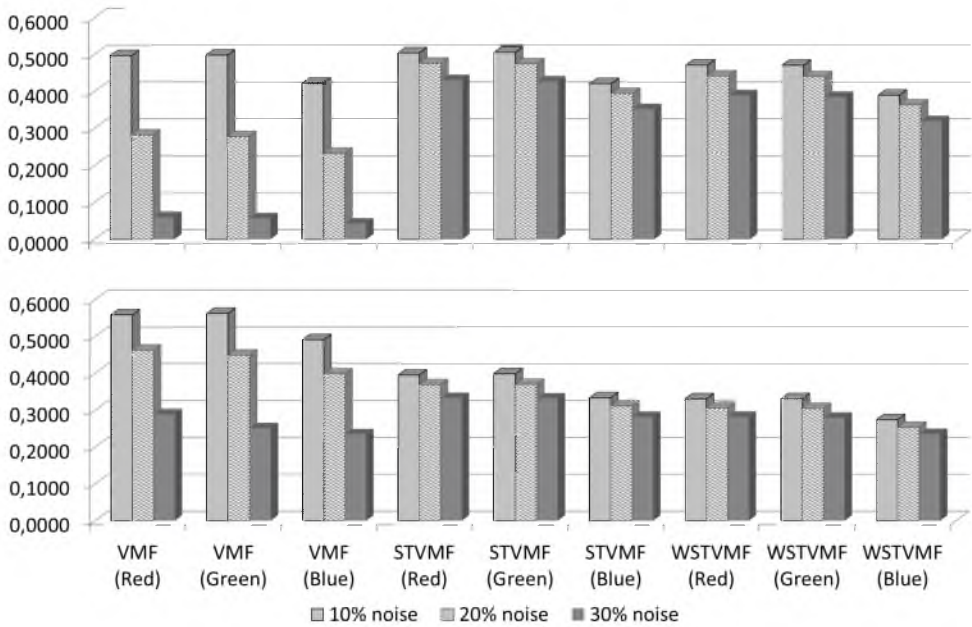


Fig. 13. CQM values for the upper and lower parts of the frames

It can be easily noticed that the static parts of the frame should not be filtered using the standard VMF filter in the presence of strong noise. A serious loss of quality can be preserved using the spatio-temporal filtering, especially for the areas with low variance. Lower parts of the frames containing many moving details can achieve better quality using the VMF filter for small amount of noise.

The application of the weighting spatio-temporal filter does not lead to significantly better results which could justify the increased computational complexity. Many image quality metrics indicate even worse quality of such filtered images in comparison to unweighted spatio-temporal VMF.

The usage of the R-SVD metric for the assessment of filtering results is at least doubtful, since the metric indicate the increase of the quality for the images containing more unfiltered noisy pixels. It is worth to notice that the higher values of this metric are related to lower quality, similarly as for the MSE.

5 Conclusions

In the future research the application of an adaptive weighted spatio-temporal filter utilizing also the motion vectors should reduce the motion blur visible in the lower part of each frame and further improve the overall image quality. Since many authors still use the MSE and similar metrics for the assessment of the filtering results, the development of the new filters optimised in the aspect of some modern image quality assessment methods seems to be an interesting area of further research. Such approach should result in much better consistency of obtained filtering results with human perception.

Presented results may also lead to the development of the hybrid adaptive filter, which should estimate the local amount of noise in order to use the standard VMF filtering procedure for less noisy areas and its spatio-temporal version for more noisy fragments of each frame.

References

- [1] Astola J., Haavisto P., Neuvo Y.: Vector Median Filters. *Proceedings of IEEE*, vol. 78 no. 4 (1990), pp. 678–689.
- [2] Bartkowiak M., Domański M.: Vector median filters for processing of color images in various colour spaces. *Proceedings of the 5-th International Conference on Image Processing and its Applications*, Edinburgh, UK, 1995, pp. 833–836.
- [3] Louverdis G., Vardavoulia M., Andreadis I., Tsalides Ph.: A new approach to morphological color image processing. *Pattern Recognition*, vol. 35 (2002), pp. 1733–1741.
- [4] Mansouri A., Mahmoudi-Aznavah A., Torkamani-Azar F., Jahanshahi J.: Image Quality Assessment Using the Singular Value Decomposition Theorem. *Optical Review*, vol. 16 no. 2 (2009), pp. 49–53.

- [5] Okarma K.: Combined Full-Reference Image Quality Metric Linearly Correlated with Subjective Assessment. *Lecture Notes in Artificial Intelligence*, vol. 6113 (2010), pp. 539–546.
- [6] Okarma K.: Video Quality Assessment Using the Combined Full-Reference Approach. *Advances in Intelligent and Soft Computing*, 2010, vol. 84 (Image Processing and Communications Challenges 2), pp. 51–58.
- [7] Okarma K.: Colour Image Quality Assessment Using the Combined Full-Reference Approach. *Advances in Intelligent and Soft Computing*, 2011, vol. 95 (Computer Recognition Systems 4), pp. 287–296.
- [8] Ponomarenko N., Lukin V., Zelensky A., Egiazarian K., Carli M., Battisti F.: TID2008 - a Database for Evaluation of Full-Reference Visual Quality Assessment Metrics. *Advances of Modern Radioelectronics*, vol. 10 (2009), pp. 30–45.
- [9] Sheikh H., Bovik A.: Image Information and Visual Quality. *IEEE Transactions on Image Processing*, vol. 15 no. 2 (2006), pp. 430–444.
- [10] Vardavoulia M., Andreadis A., Tsalides Ph.: A New Vector Median Filter for Colour Image Processing. *Pattern Recognition Letters*, vol. 22 (2001), pp. 675–689.
- [11] Wang Z., Bovik A., Sheikh H., Simoncelli E.: Image Quality Assessment: From Error Measurement to Structural Similarity. *IEEE Transactions on Image Processing*, vol. 13 no. 4 (2004), pp. 600–612.
- [12] Wang Z., Simoncelli E., Bovik A.: Multi-Scale Structural Similarity for Image Quality Assessment. *Proc. 37th IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, 2003, pp. 1398–1402.
- [13] Zhang L., Zhang L., Mou X.: RFSIM: A Feature Based Image Quality Assessment Metric Using Riesz Transforms. *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, Hong Kong, China, 2010, pp. 321–324.
- [14] Zhang L., Zhang L., Mou X., Zhang D.: FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing*, vol. 20 no 8 (2011), pp. 2378–2386.