https://doi.org/10.2352/ISSN.2470-1173.2021.10.IPAS-237

Decision-making on image denoising expedience

Andrii Rubel, National Aerospace University, 61070, Kharkiv, Ukraine
Oleksii Rubel, National Aerospace University, 61070, Kharkiv, Ukraine
Vladimir Lukin, National Aerospace University, 61070, Kharkiv, Ukraine
Karen Egiazarian, Tampere University, FIN 33101, Tampere, Finland

Abstract

Image denoising is a classical preprocessing stage used to enhance images. However, it is well known that there are many practical cases where different image denoising methods produce images with inappropriate visual quality, which makes an application of image denoising useless. Because of this, it is desirable to detect such cases in advance and decide how expedient is image denoising (filtering). This problem for the case of well-known BM3D denoiser is analyzed in this paper. We propose an algorithm of decision-making on image denoising expedience for images corrupted by additive white Gaussian noise (AWGN). An algorithm of prediction of subjective image visual quality scores for denoised images using a trained artificial neural network is proposed as well. It is shown that this prediction is fast and accurate.

Introduction

One of the main obstacles in image processing applications significantly affecting image perceptual quality is a presence of noise [1]. Due to this, image denoising is an important step in image preprocessing to suppress noise and improve image visual quality. However, the use of existing image denoising methods can often lead to a degradation of image visual quality resulting in loss of important details, edges and texture features [2-4]. Such distortions, appearing after applying denoising, influence efficiency of subsequent high-level image processing tasks, e.g. semantic segmentation [5]. Often consumers are unsatisfied by denoising results because of introduced visually annoying structural distortions. Hence, monitoring and predicting an appropriate perceptual quality of denoised images is extremely important to meet required quality of experience (QoE) for end-users [6].

Generally, image quality can be evaluated in two ways – objectively, by image quality measures, and subjectively, by involving humans to perform visual quality assessment [1, 3, 6]. Since large scale subjective evaluations are very complicated and cannot be used for real-time image quality assessment, image quality measures (metrics) are widely used to assess quality of denoised images. Image denoising efficiency is typically described by improvement of image quality measures, i.e. difference between the measure values after and before denoising [7]. At a first glance, it seems that a high value of improvement for a given measure always corresponds to an image enhancement. Nevertheless, there are situations where visual quality of denoised image has not been considerably improved compared to the input noisy image regardless a positive value of improvement of a given measure [4, 7]. In that case, a positive effect of denoising is negligible and it is not worth applying denoising operation. According to this, it is highly desirable to detect such situations far in advance and to undertake a decision on image denoising expedience – is it worth applying denoising for a given image or it is better to skip it and thus save a processing time.

Although many image visual quality measures taking into account peculiarities of human vision system have been developed, subjective evaluation is still the most adequate assessment of image visual quality [1, 6].

Over the last years, several studies concerning subjective evaluation of visual quality of denoised images have been conducted [3, 8, 9]. In [8], subjective evaluations have been conducted to assess visual quality for images processed by BM3D with four different hard thresholding parameters for a single noise intensity. In [9], a special database SubjectiveIQA has been introduced and the subjective evaluation experiments for images denoised by different filters have been done. It has been shown that denoising is mostly reasonable to apply if images are corrupted by a moderate intensity noise and if images having simple content, whilst, for highly textured images, denoising often leads to visible degradation.

In this study, we analyze denoising expedience for images corrupted by additive white Gaussian noise (AWGN) for the case of BM3D denoiser [10] using full-reference image quality measures. We conduct our analysis based on SubjectiveIQA database with the provided subjective visual quality scores. Finally, we propose an algorithm for decision-making on image denoising expedience.

Analysis of denoised images using full-reference image quality measures

Before carrying out our detailed analysis of visual quality for denoised images, let us briefly describe the SubjectiveIQA database which was introduced in [9] for analysis of visual quality of denoised images.

The SubjectiveIQA database consists of 16 reference grayscale images, 112 images corrupted by AWGN with seven noise levels (noise standard deviations used are equal to 3, 5, 10, 15, 20, 25 and 30, respectively) and 224 denoised images, where each noisy image has been denoising by both DCT-filter and BM3D filter [9]. A methodology of subjective evaluation was the pairwise comparison. Each participant had to choose an image with better perceptual quality – noisy image or a denoised one. Subjective scores in the database are presented as probabilities of voting for denoising P_{vote}, meaning that the denoised image has a better perceptual quality than the corresponding noisy one. Thus, a probability P_{vote} tending to unity corresponds to the case when practically all participants prefer a denoised image over a noisy one. Probability P_{vote} greater than 0.5 means that the use of denoising is, probably, expedient, while if P_{vote} is less than 0.5 then it is not worth applying denoising for a given image [9].

Fig. 1 shows an example of image fragment for which the use of denoising is obviously expedient, since the obtained probability of voting for denoising is about 0.9 in the case of noise standard deviation (STD) equal to 10. From the presented denoised fragment, it is well seen that BM3D demonstrates good noise suppression and edge preservation.

Fig. 2 shows an example of image fragment for which applying denoising is clearly useless. In this case, the obtained probability P_{vote} is less than 0.4 for noise STD equal to 10. The fragment of denoised image looks very similar to the noisy one. In this example, noise is visually masked by a texture.

***Figure 1**. Examples of image fragments (reference image – left, noisy image corrupted by AWGN with STD=10 – middle, denoised - right)*

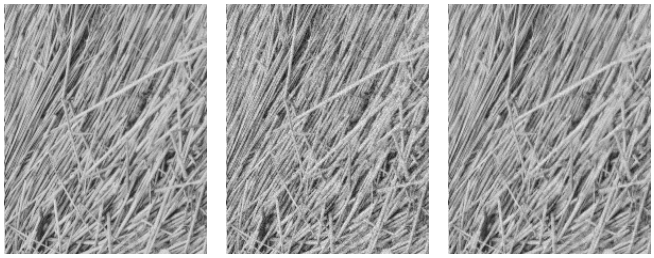


***Figure 2**. Examples of texture image fragments (reference image – left, noisy image corrupted by AWGN with STD=10 – middle, denoised - right)*

Given that a huge number of image quality assessments have been proposed during the last decades, it is necessary to find out which of them are the most appropriate and adequate to characterize visual quality of denoised images. To meet this need, we have evaluated performance of more than 40 full-reference image quality measures. Performance evaluation has been carried out by calculating Spearman rank order correlation coefficient (SROCC) between the values of improvements of the considered image quality measures and subjective scores, i.e. probabilities of voting P_{vote} for denoised images from the SubjectiveIQA database. Results of the conducted evaluation are shown in Fig. 3.

From the results presented in Fig. 3, it can be observed that none of the considered image quality measures provides SROCC value higher than 0.82. Most of the considered full-reference image quality measures demonstrate moderate correspondence to subjective visual quality scores. The commonly used peak-signal-to-noise ratio (PSNR) measure has SROCC value about 0.6. Among the best performing image quality measures, there are the well-known FSIM measure [11] and recently proposed image quality measures ADD-SSIM [12] and SSIM4 [13]. The best SROCC value is 0.81.

Let us analyze best performing image quality measures in more details. Scatterplots of the probabilities of voting for denoising P_{vote} and the values of improvements for image quality measures FSIM, SSIM4, ADD-SSIM and PSNR-HVS-M [14] are shown in Fig. 4 and Fig. 5. In the presented scatterplots, an asterisk indicates images with rather simple structure, while a square represents data for highly textured images. Note that the improvements for most measures can be quite easily and accurately predicted [4, 7].

First, it is well seen that the obtained scatterplots for quality measures FSIM (see Fig.4), SSIM4 and ADD-SSIM (see Fig. 5) are almost identical. Therefore, we will further analyze the values of improvement for FSIM because it is widely used and well-studied.

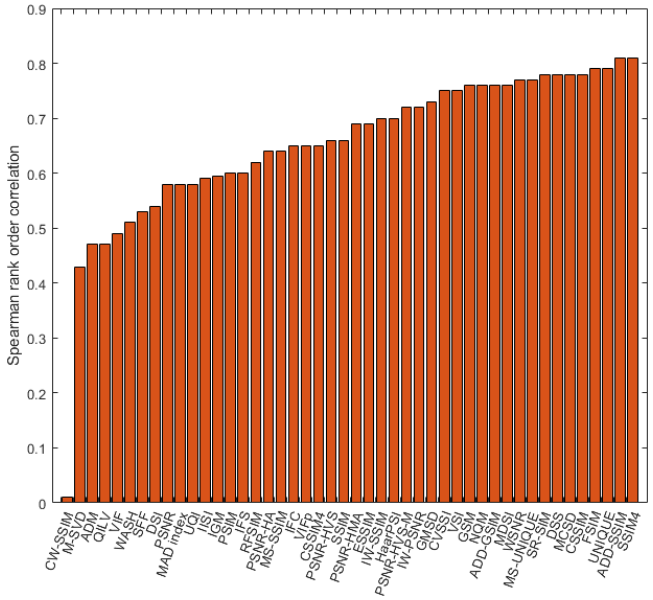


***Figure 3**. SROCC values between improvements of the considered image quality measure and subjective scores*

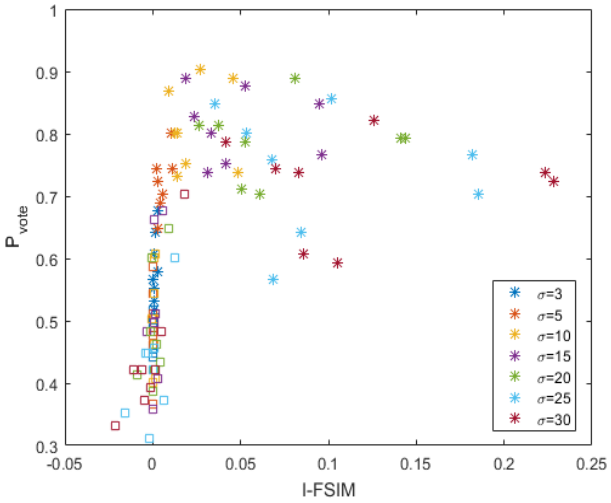
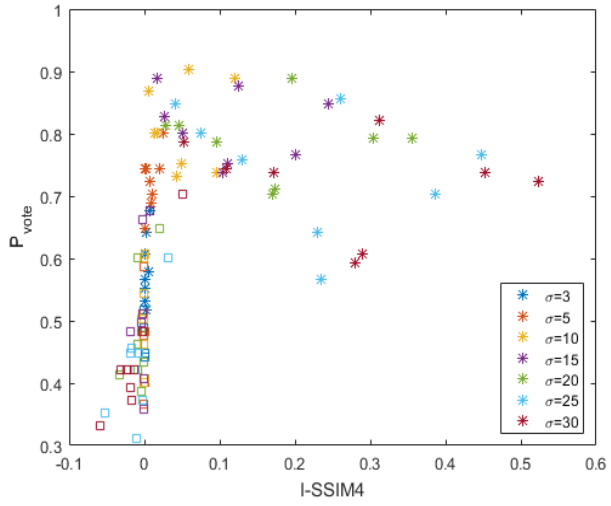
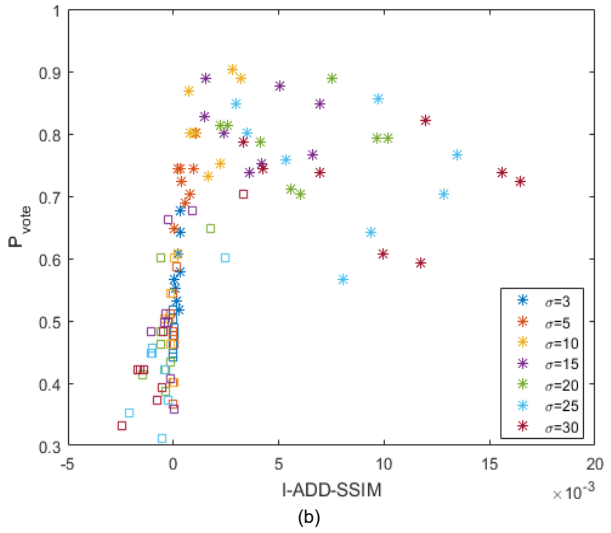


***Figure 4**. Scatterplot of probabilities of voting for denoising P_{vote} and improvements of FSIM*

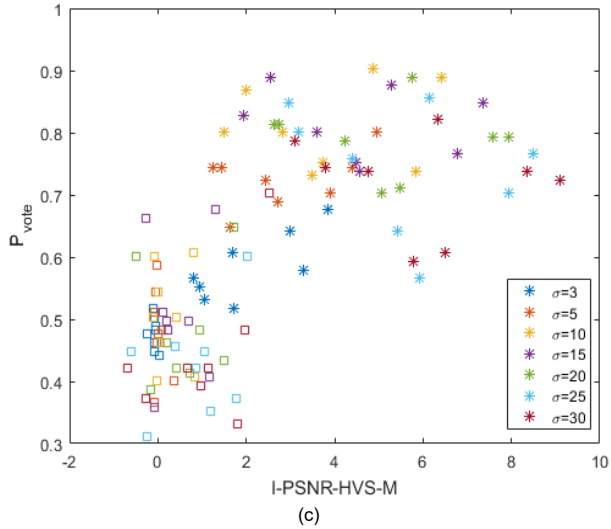
From the analysis of the scatterplot of values of FSIM improvement (denoted as I-FSIM) and the probabilities of voting for denoising P_{vote}, it becomes clear that if the values of improvement I-FSIM are negative (i.e. the value of the measure FSIM after applying denoising has become less than for the original noisy image), then denoising is useless. In this case, the values of P_{vote} do not exceed 0.4. On the other hand, when the value of improvement I-FSIM is greater than 0.01, one can state that applying the denoising is expedient (almost for all such images the probabilities exceed 0.6). At the same time, there are many images for which the values of improvements do not exceed 0.01 and the probabilities of voting vary in a wide range.



(a)



(b)



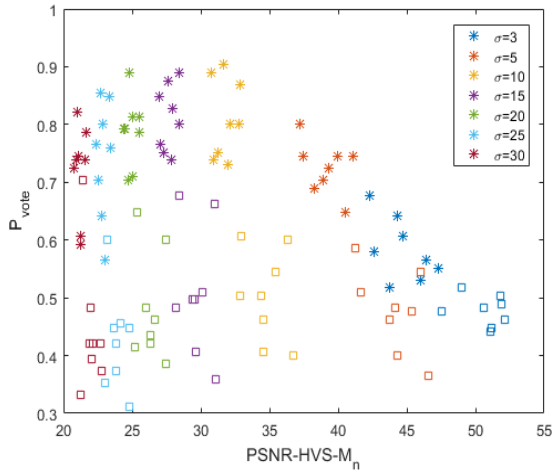
(c)

***Figure 5**. Scatterplot of probabilities of voting for denoising P_{vote} and improvements of SSIM4 (a), ADD-SSIM (b) and PSNR-HVS-M (c)*

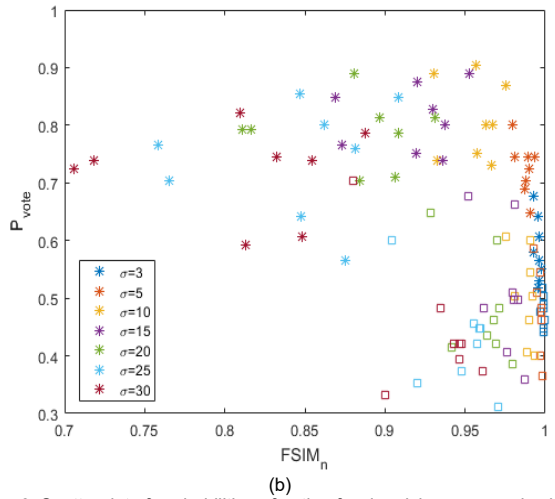
Considering that the range of possible values for FSIM, SSIM4 and ADD-SSIM measures is from 0 to 1 (larger values correspond to better image quality), it would be helpful to analyze image quality measure with other properties. For this purpose, we have chosen PSNR-HVS-M measure (SROCC value in the performance evaluation is 0.72). Recall that PSNR-HVS-M is expressed in dB and higher values indicate better image visual quality [14].

Analysis of the scatterplot obtained for the improvement of PSNR-HVS-M (see Fig. 5c) shows that the values of improvement for this measure (denoted as I-PSNR-HVS-M) greater than 2dB practically guarantee that image denoising (applying BM3D) is expedient. However, there are still many images, mostly highly textured, for which the probability of voting for denoising greatly vary depending on noise intensity and texture characteristics.

In order to understand when image denoising is expedient, it is also necessary to analyze the values of quality measures of the original noisy images. This is explained by the fact that noise in images may be not visually noticeable or practically invisible; and, therefore, it makes denoising useless [15]. To do this, we use PSNR-HVS-M and FSIM measures. Examples of scatterplots of probabilities of voting and values for PSNR-HVS-M and FSIM measures calculated for original noisy images are shown in Fig. 6.



(a)



(b)

***Figure 6**. Scatterplot of probabilities of voting for denoising P_{vote} and values of PSNR-HVS-M (a) and FSIM (b) measures calculated for noisy images*

Various studies have shown that PSNR-HVS-M and FSIM measures adequately characterizes visual quality for noisy images [3]. From the obtained dependences (see Fig. 6), it is worth noting the following. First, for noisy images for which values of FSIM are about 0.97-0.99, denoising is not expedient. Second, for noisy images with FSIM values less than 0.95, except highly textured images, denoising results in image enhancement and it is expedient. For images with PSNR-HVS-M greater than 45 dB, a denoising is useless.

Moreover, let us analyze dependences between values of improvements of PSNR-HVS-M and FSIM image quality measures and noise intensity (STD). Fig. 7 shows the scatterplots of values of improvement for different measures and noise intensities.

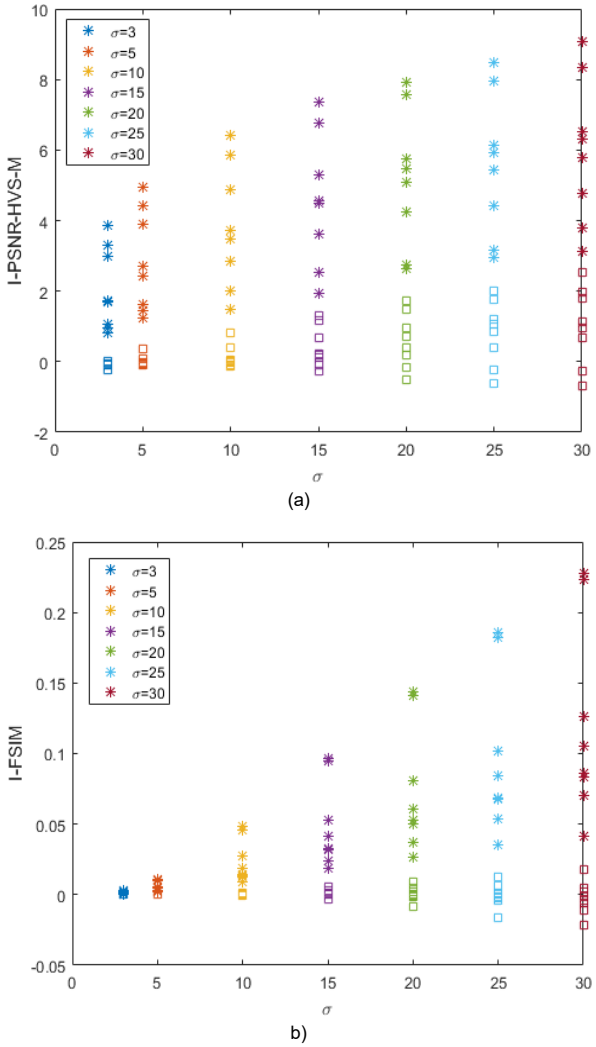


***Figure 7**. Scatterplot of improvements for PSNR-HVS-M (a) and FSIM(b) measures and AWGN standard deviations*

The scatterplot of values of improvement for PSNR-HVS-M measure is especially interesting (see Fig. 7a). In this case, one can clear identify images for which denoising is useless: the values of I-PSNR-HVS-M do not exceed 2 dB and the probabilities of voting for denoising are mostly less than 0.5. Thus, it is possible to establish such a dependence, if I-PSNR-HVS-M $> 0.08 \cdot \sigma_n$, then applying denoising is expedient, otherwise it is useless.

Based on the aforementioned observation and remarks, we propose an algorithm on image denoising expedience, which is shown in Fig. 8.

Let us verify the proposed algorithm and compare the results with the probabilities of voting for denoising of images from SubjectiveIQA database. Note that we will consider the use of denoising expedient if subjective scores (probabilities of voting for denoising P_{vote} are greater than 0.6).

The correctness of the decision-making according to the proposed algorithm for images denoised by the BM3D filter is 92.86% which ensures a reliable decision.

An example of an image for which the algorithm incorrectly indicates that it is not useful to apply BM3D for noise STD equal to 10, 15, 20 is shown in Fig. 9. This is due to the ability of texture to mask a noise.

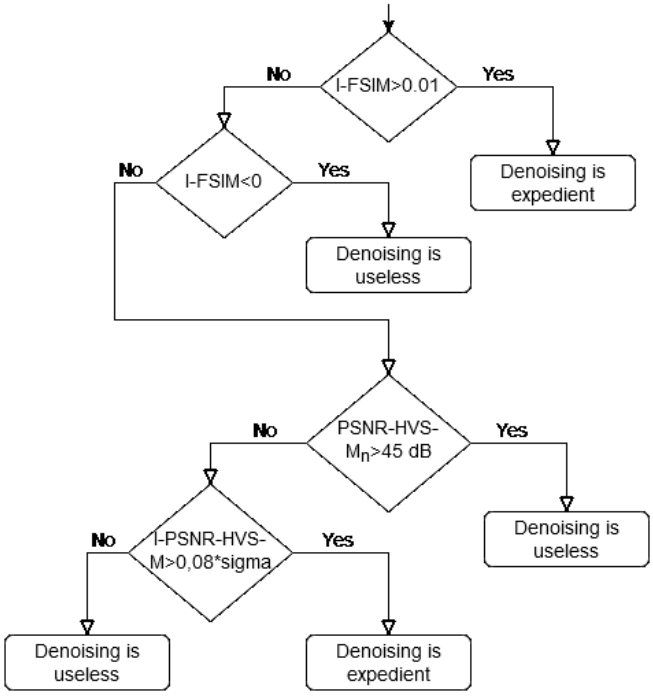


***Figure 8.** Algorithm for decision-making on image denoising expedience*

***Figure 9.** Example of an image for which the algorithm incorrectly indicates the usefulness of applying denoising*

Prediction of subjective scores for denoised images

To undertake a decision on image denoising expedience let us consider a problem of predicting subjective quality scores for denoised images. To this end, artificial neural network is as a predictor. Values of improvements of the most appropriate image quality measures in accordance with the values of Spearman's rank order correlation coefficient (SROCC) calculated between the improvements of image quality measures and subjective scores P_{vote} for the SubjectiveIQA database (available at: https://github.com/ViA-RiVaL/SubjectiveIQA) are used as inputs.

We have employed a feedforward multilayer back propagation neural network to predict subjective image quality scores (represented in the SubjectiveIQA database by P_{vote}). The neural network consists of an input layer, two hidden layers of 3 and 2 neurons, respectively, and an output layer. For each of the hidden layers, hyperbolic tangent activation function is used, while the output layer is activated by a linear function. The architecture of the employed neural network is shown in Fig. 10. The inputs of the neural network are values of improvements of full-reference image quality measures, namely I-FSIM, I-SSIM4 and I-ADD-SSIM. The neural network architecture has been empirically selected during preliminary training and validation experiments.

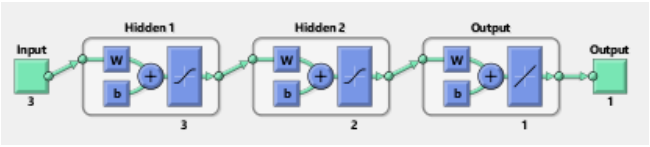


Figure 10. *Architecture of a neural network*

To assess a prediction accuracy of the trained neural network, goodness-of-fit adjusted R^2 (coefficient of determination) along with root mean square error (RMSE) are used. A coefficient of determination R^2 is the square of the linear correlation between the predicted values and ground truth values [16]. It ranges from 0 to 1, where high values of R^2 correspond to better prediction accuracy. At the same time, lower the value of RMSE, more accurate the prediction is.

Since the number of training epochs affects a prediction accuracy at the testing (validation) stage, we have conducted additional intensive experiments. In this regard, the neural network has been trained on different number of epochs varied from 15 to 25 with a step 1. Various training/testing dataset split ratios (namely 90/10%, 80/20%, 70/30% and 60/40%) have been also studied. To obtain more reliable results of a prediction accuracy, the procedure of random splitting into training and test sets has been carried out 1000 times for each of the considered number of epochs and split ratios. Training of neural network took about 0.6 sec. All experiments were conducted on a computer with Intel Core i7-2670 QM processor and 16 Gb of RAM. The obtained dependences are shown in Fig. 11.

Analysis of these dependences shows the following. It is well seen that the best prediction accuracy in terms of coefficient of determination (adjusted R^2) is provided for the training/testing dataset split ratio of 70/30 % and 25 training epochs. Second, the optimal number of training epochs for the training/testing dataset split ratios of 80/20% and 90/10 % is 24, while for the split ratio of 60/40 % is 23.

Summarized results of RMSE and adjusted R^2 obtained by averaging across 1000 random splitting into train/test sets and their standard deviations for different splitting ratios and optimal number of training epochs are presented in Table 1.

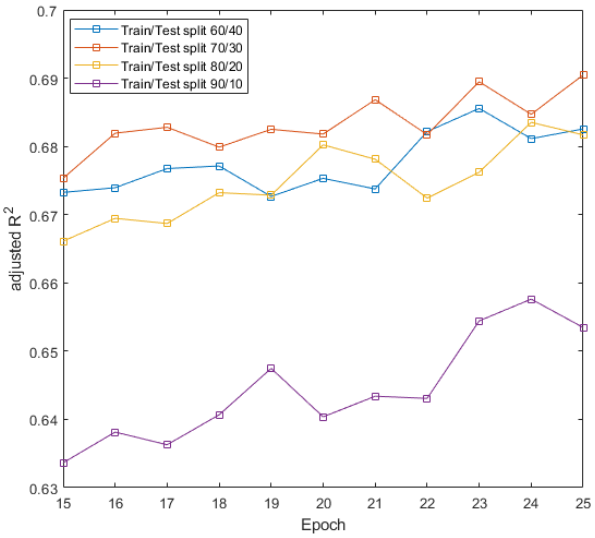


Figure 11. *Dependences of adjusted R^2 on different number of training epochs for various training/testing dataset split ratios*

Table 1: Prediction Accuracy

Train/Test dataset split ratio	RMSE	STD of RMSE	Adjusted R^2	STD of adjusted R^2
60/40	0.0870	0.0115	0.6855	0.0853
70/30	0.0852	0.0117	0.6905	0.088
80/20	0.0829	0.0139	0.6835	0.1105
90/10	0.0794	0.018	0.6576	0.1491

As can be seen from Table 1, the best prediction accuracy in terms of RMSE is provided for the training/test dataset split ratio of 90/10% which is normal since the neural network is trained on a larger amount of data. Meanwhile, for this case, the STD value of RMSE is greater than for the cases of other splitting ratios that influences stability of the prediction results. For all the cases of splitting ratios, values of RMSE do not exceed 0.09. Values of STD of RMSE and adjusted R^2 are high, which can be explained by the limited size of the dataset.

To further improve a prediction accuracy and obtain more reliable results, it is desirable to expand the existing database of denoised images by adding more test images, noise levels and obtaining more subjective scores. Thus, a prediction is quite accurate, but, in future work, we will study more accurate prediction models of subjective scores for denoised images.

Conclusions

In this paper, analysis of visual quality for denoised images using appropriate full-reference image quality assessment measures along with the provided subjective image quality scores and image dataset has been carried out. Based on this analysis, the algorithm for decision-making on image denoising expedience for images corrupted by AWGN in the form of a sequence of comparisons to the thresholds and logic operations has been proposed. Its testing has shown that it is possible to make a reliable decision on image denoising expedience in advance without carrying out denoising. Additionally, a method of subjective image quality scores prediction for images denoised by the BM3D filter using artificial neural network has been proposed. It has been shown that the subjective image quality scores can be predicted rather accurately with goodness-of-fit R-squared about 0.69.

References

[1] D. M. Chandler, "Seven challenges in image quality assessment: Past present and future research", *ISRN Signal Process.*, vol. 2013, Nov. 2013.

[2] A. Pižurica, Image Denoising Algorithms: From Wavelet Shrinkage to Nonlocal Collaborative Filtering, Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1-17, 2017.

[3] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," Signal Processing: Image Communication, vol. 30, pp. 57-77, Jan. 2015, doi: 10.1016/j.image.2014.10.009.

[4] O. Rubel, V. Lukin, S. Abramov, B. Vozel, O. Pogrebnyak, and K. Egiazarian, "Is Texture Denoising Efficiency Predictable?," International Journal of Pattern Recognition and Artificial Intelligence, vol. 32, no. 01, p. 32, 2018, doi: 10.1142/S0218001418600054.

[5] D. Liu, B. Wen, X. Liu, Z. Wang and T. Huang, "When Image Denoising Meets High-Level Vision Tasks: A Deep Learning Approach", *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018, p. 842-848, doi:10.24963/ijcai.2018/117.

[6] Z. Wang and A. C. Bovik, Modern Image Quality Assessment, San Mateo, CA, USA:Morgan Claypool, 2006.

[7] O. Rubel, A. Rubel, V. Lukin and K. Egiazarian, "Blind DCT-based prediction of image denoising efficiency using neural networks," in Proc. of *2018 7th European Workshop on Visual Information Processing (EUVIP)*, Tampere, 2018, pp. 1-6, doi: 10.1109/EUVIP.2018.8611710.

[8] K. Egiazarian, M. Ponomarenko, V. Lukin and O. Ieremeiev, "Statistical Evaluation of Visual Quality Metrics for Image Denoising," in Proc. of *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, 2018, pp. 6752-6756, doi: 10.1109/ICASSP.2018.8462294.

[9] A. Rubel, O. Rubel and V. Lukin, "Analysis of visual quality for denoised images," in Proc. of *2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, Lviv, 2017, pp. 92-96, doi: 10.1109/CADSM.2017.7916093.

[10] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," in *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080-2095, Aug. 2007, doi: 10.1109/TIP.2007.901238.

[11] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," in *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378-2386, Aug. 2011, doi: 10.1109/TIP.2011.2109730.

[12] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang and W. Zhang, "Analysis of Distortion Distribution for Pooling in Image Quality Prediction," in *IEEE Transactions on Broadcasting*, vol. 62, no. 2, pp. 446-456, June 2016, doi: 10.1109/TBC.2015.2511624.

[13] M. Ponomarenko, K. Egiazarian, V. Lukin and V. Abramova, "Structural Similarity Index with Predictability of Image Blocks," *2018 IEEE 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET)*, Kiev, 2018, pp. 115-118, doi: 10.1109/MMET.2018.8460285.

[14] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in Proc. 3rd Int. Workshop Video Process. Qual. Metrics Consum. Electron, Scottsdale, USA, Jan. 2007, 4 p.

[15] N. Ponomarenko, V. Lukin, J. Astola and K. Egiazarian, "Analysis of HVS-Metrics' Properties Using Color Image Database TID2013", *Proc. of Int. Conf. ACIVS 2015*, pp. 613-624, 2015, doi: 10.1007/978-3-319-25903-1_53.

[16] A.C. Cameron and F. Windmeijer, "An R-squared measure of goodness of fit for some common nonlinear regression models," Journal of Econometrics, vol. 77, no. 2, pp. 329-342, April 1997.

Author Biography

Andrii Rubel received his M.S. in telecommunications from the Kharkiv National University of Radioelectronics, in 2016. Since 2016 he is a Ph.D. student at the National Aerospace University named after M.E. Zhukovsky. His research interests include image processing, image quality assessment and machine learning.

Oleksii Rubel received his M.S. in telecommunications from the National Aerospace University named after M.E. Zhukovsky, in 2013. He received his Ph.D. in Remote Sensing from the National Aerospace University named after M.E. Zhukovsky, in 2016. His research interests include image enhancement, image and video processing and deep learning.

Vladimir Lukin received his Ph.D. in 1988 and Doctor of Technical Science in remote sensing from the National Aerospace University named after M.E. Zhukovsky, Ukraine, in 2002. Since 1995, he has been in cooperation with Tampere University of Technology. Currently he is a head of the Department of Information and Communication Technologies and a professor. His research interests include digital signal/image processing, remote sensing data processing, image denoising and compression.

Karen Egiazarian received his Ph.D. from Moscow M. V. Lomonosov State University, Russia, in 1986, and his Doctor of Technology in signal processing from Tampere University of Technology, Finland, in 1994. He is leading a "Computational imaging" group and a professor at the Department of Computing Sciences, Tampere University, Tampere, Finland. He is an IEEE Fellow, and a member of the DSP Technical Committee of the IEEE Circuits and Systems Society. During 2016-2020 he has been Editor-in-Chief of the Journal of Electronic Imaging. His main research interests are in the field of computational imaging, compressed sensing, efficient signal processing algorithms, image/video restoration and compression.

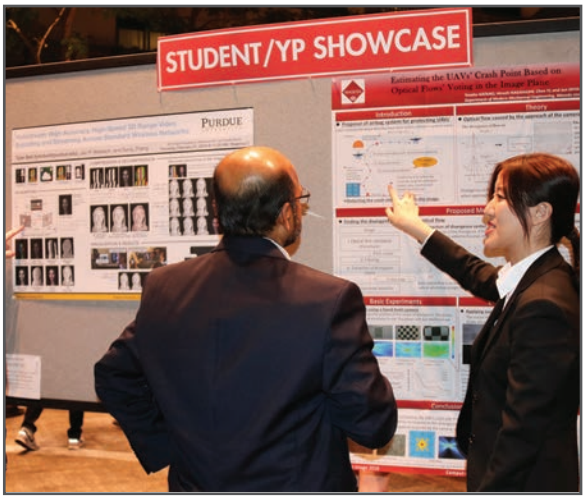
JOIN US AT THE NEXT EI!

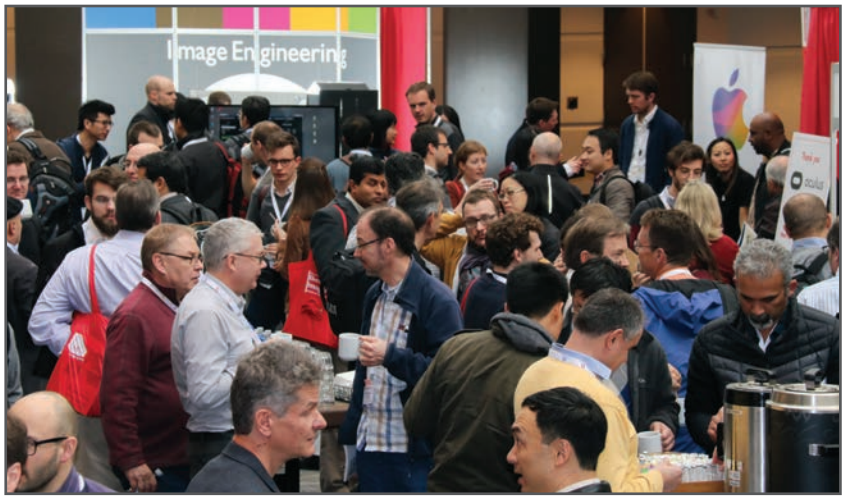
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org