Current Trends and Advances in Image Quality Assessment

Krzysztof Okarma

Department of Signal Processing and Multimedia Engineering, Faculty of Electrical Engineering,
West Pomeranian University of Technology, Szczecin,
Sikorskiego 37, 70-313 Szczecin, Poland
okarma@zut.edu.pl

Abstract—Image quality assessment (IQA) is one of the constantly active areas of research in computer vision. Starting from the idea of Universal Image Quality Index (UIQI), followed by well-known Structural Similarity (SSIM) and its numerous extensions and modifications, through Feature Similarity (FSIM) towards combined metrics using the multi-metric fusion approach, the development of image quality assessment is still in progress. Nevertheless, regardless of new databases and the potential use of deep learning methods, some challenges remain still up to date. Some of the IQA metrics can also be used efficiently for alternative purposes, such as texture similarity estimation, quality evaluation of 3D images and 3D printed surfaces as well as video quality assessment.

Index Terms—Image analysis; Image quality assessment.

I. INTRODUCTION

Development of computer vision and image analysis methods and growing variety of their applications cause the increasing interest in accuracy of results obtained using such algorithms. In many applications the results of image analysis may be used to make some decisions and therefore they should be as reliable as possible. Nevertheless, most of computer vision algorithms require high quality input images to ensure their proper work and therefore the knowledge of image quality assessment (IQA) results is strongly desired. For example, analysing the consecutive frames of a video stream, some of them could be skipped, if low image quality is detected, to prevent their influence on the results of analysis. Such a strategy can be useful e.g. for machine vision control of mobile robots.

The most widely known classical IQA methods rely on the comparison of respective pixels of two images with the assumption of the same resolution and a perfect adjustment of both images – one of them is considered as a "pristine" reference image and the second one is subjected to some distortions or contains some contaminations.

Such approach, based on the comparison with the reference image, is referred as full-reference (FR) assessment, whereas "blind" methods, which do not demand the knowledge of the original image, are known as referenceless or no-reference (NR) ones. In some cases, the partial information related to the original image is known and therefore the reduced-reference (RR) methods can also be applied, assuming its availability.

Considering numerous different IQA methods proposed during recent years, it can be stated that FR metrics are more universal, although in many applications where the reference image is unavailable, e.g. live video transmission, "blind" methods are much more desired.

II. FULL-REFERENCE IMAGE QUALITY ASSESSMENT

A. Structural Similarity and its Modifications

Since the classical pixel based IQA metrics, such as Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR), are poorly correlated with perception of various types of distortions by human observers, some other approaches have been proposed which are additionally more robust to small mutual translations of both compared images. The first such idea has been the Universal Image Quality Index [1] based on the product of three factors corresponding to luminance distortions, loss of contrast and structural distortions calculated using the sliding window approach. Obtained quality map is further averaged leading to the final metric. As for the dark or flat areas of images (with constant brightness) the results may become unstable, due to the presence of zero variance in the denominators, the further modification of this approach has led to probably the most popular Structural Similarity (SSIM) metric [2], containing additional small constants preventing the division by zero. Its formula for calculations of the local similarity factors can be defined as

$$SSIM(x, y) = \frac{(2\overline{xy} + C_1) \times (2\sigma_{xy} + C_2)}{(\overline{x}^2 + \overline{y}^2 + C_1) \times (\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (1)$$

where x and y are the 11×11 pixels local fragments of assessed and reference images weighted with 2D Gaussian window with small constants preventing the division by zero: $C_1 = (0.01 \times L)^2$ and $C_2 = (0.03 \times L)^2$ assuming $L = 256$ being the number of luminance levels.

Over the next several years this approach has been often modified being the basis for numerous extended metrics. The first, proposed by the creators of the SSIM, is its multi-scale version referred as MS-SSIM [3] expressed as

$$MS - SSIM(x, y) =$$
$$= [l(x, y)]^{\alpha_M} \times \prod_{j=1}^{M} [c(x, y)]^{\beta_j} \times [s(x, y)]^{\gamma_j}, \quad (2)$$

where the contrast and structural factors are weighted for each scale and the luminance is calculated only for full resolution image with the weights optimized by its authors.

Some other extensions of the SSIM idea have been proposed independently, such as gradient based SSIM [4], the use of local variance (QILV) with the same formula [5], Gradient Similarity [6], three-component SSIM [7] or the simplified fast SSIM [8]. Some mathematical properties of the SSIM have been analysed by Brunet *et al.* [9], whereas its reduced-reference estimation has been proposed by Rehman and Wang [10] and its quaternion version for colour images by Kolaman and Yadid-Pecht [11].

Similar formulas can also be applied for the images converted to some other domains e.g. using the wavelet transform. Some exemplary metrics have been provided as well, e.g. Complex Wavelet SSIM (CW-SSIM) [12], Information Weighting SSIM (IW-SSIM) [13] and Edge Strength Similarity (ESSIM) [14].

An interesting modification of this general idea has been proposed with the use of Riesz transform leading to RFSIM metric [15], being the basis for widely accepted Feature Similarity (FSIM) metric [16] defined as

$$FSIM = \frac{\sum_u \sum_v S(u,v) \times PC_{max}(u,v)}{\sum_u \sum_v PC_{max}(u,v)}, \qquad (3)$$

with phase congruency (PC) and gradient magnitude (G) factors calculated according to the SSIM-like formula

$$S(u,v) = \left(\frac{2 \times PC_1(u,v) \times PC_2(u,v) + T_{PC}}{PC_1^2(u,v) + PC_2^2(u,v) + T_{PC}} \right)^{\alpha} \times$$

$$\times \left(\frac{2 \times G_1(u,v) \times G_2(u,v) + T_G}{G_1^2(u,v) + G_2^2(u,v) + T_G} \right)^{\beta}, \qquad (4)$$

where PC_{max} is the higher from two local values of phase congruency from the reference and assessed image, *(u,v)* is the position of the sliding window in images 1 and 2, and *T* prevents the division by zero. Gradient values can be obtained using one of the popular convolution filters (Scharr mask is recommended by the FSIM authors). Its colour version, known as FSIMc, contains the additional factor (with the exponent γ), where the same formulas are applied for both chrominance components in YIQ colour space.

Another direction of research related to the application of the idea of the Structural Similarity is its use for texture similarity, leading to the idea of STSIM metric [17], useful e.g. for Content Based Image Retrieval (CBIR) purposes.

B. Some Other Full-Reference IQA Metrics

Regardless, of the popularity of Structural Similarity and numerous SSIM based methods, some alternative methods have been proposed by various researchers. Some of the representative trends include the use of information theory, e.g. Information Fidelity Criterion (IFC) [18], further extended into Visual Information Fidelity (VIF) [19], as well as Human Visual System (HVS) based models applied e.g. in NQM [20] and VSNR [21] metrics.

Some recently proposed approaches incorporating the similarity assessment for IQA purposes are the Haar wavelet-based perceptual similarity index (HaarPSI) [22] and SuperPixel-based similarity index (SPSIM) [23].

One of the most promising alternative approaches to the FR IQA seems to be the application of Singular Value Decomposition (SVD). The first approach to its use for this purpose, namely M-SVD [24], utilizes the Euclidean distances between the vectors of singular values obtained for 8×8 pixels blocks. Nonetheless, in this method only the singular values have been used without the analysis of the remaining coefficients and it has been verified for only 5 test images subjected to six types of distortions.

The two other interesting methods have been presented in 2009. The first one is the Reflection Factor (RF) [25], which utilizes not only the singular values but also right singular vector matrices to calculate the reflections, denoted as total weighted differences scaled by the singular values dependent on the energy of the image. The second algorithm, known as R-SVD [26], is based on the "referee matrix" obtained by the substitution of the reference image's left singular matrices by their equivalents calculated for the distorted one. These calculations have been made using the sliding window approach using the 8×8 pixels image fragments to determine the local quality. Both methods have been tested using LIVE Image Quality Assessment Database, described in Section V.

In the paper [27] SVD has been employed to separate the structural (content-dependent) and the content-independent components, further fused to obtain the final quality scores verified using the TID2008 dataset (discussed later as well). The utilization of local singular value decomposition (SVD) as a structural projection tool to select local image distortion features followed by perceptual spatial pooling and neural networks to predict a perceptual quality score has been presented by Hu *et al.* [28] and verified using 3 databases.

The most recent attempt to use the SVD for IQA purposes has been presented as the evaluation of the structural information utilizing its reflection on the original singular vector matrices, denoted as Structural SVD (SSVD) [29]. In this metric both singular values and vectors are used to effectively estimate the perceived visual quality. The overall quality metric is the combination of separately estimated structural and luminance degradations calculated for local blocks with additional regulating factor and further averaged. The performance of the SSVD metric has been evaluated using 9 state-of-the-art IQA databases. Interestingly, it demonstrates the best performance for the newest TID203 database, whereas for the older and smaller TID2008, better results can be achieved for HaarPSI.

Nevertheless, the list of the above mentioned metrics is far from completeness due to numerous different IQA metrics proposed in recent years by many researchers. Some of them are available together with the wrapper code in the MeTriX MuX package provided by Matthew Gaubatz [30].

Moreover, recent trends related to the use of deep learning methods will probably cause the development of many new algorithms, including no-reference algorithms. Another expected direction of research is the development of colour specific metrics as most of these mentioned above are designed for grayscale images only.

III. MULTI-METRIC FUSION AND COMBINED METRICS

A. Correlation with Subjective Quality Assessment

The most typical approach to the development of a new image quality metric is related to the increase of its correlation with subjective scores. To verify the appropriateness of newly proposed metrics, some IQA databases have been provided, containing numerous images subjected to various types of distortions together with Mean Opinion Score (MOS) values or Differential MOS values, obtained as a result of analysis of opinions expressed by independent human observers. The most popular publicly available IQA datasets are briefly discussed in Section V.

Although high correlation of automatically calculated metric's values with subjective scores are not always the most relevant element, in most of the applications – especially related to computer visualization – can be considered as a reasonable objective. Since most image analysis algorithms provide results strongly dependent on the quality of input images, only relatively small number of robust methods, e.g. Hough line detection, can be considered as exceptions. Nevertheless, it is worth remembering that in some machine vision applications high correlation of IQA results with subjective quality assessment may not always be of the highest priority.

Three most common approaches to calculation of the correlation between the subjective and objective scores are: Pearson Linear Correlation Coefficient (PCC), Spearman Rank Order Correlation Coefficient (SROCC) and Kendall Rank Order Correlation Coefficient (KROCC). Since the widely known PCC can be considered as a measure of quality prediction accuracy, both rank order coefficients should be treated as measures of prediction monotonicity.

Spearman ρ (rho) coefficient is a nonparametric measure of statistical dependence between the rankings of two variable assessing how well this relationship can be expressed using the monotonic function. In contrast to PCC which is calculated directly for the raw quality scores in both rankings of objective and subjective evaluations, SROCC is equal to the Pearson correlation between the rank values of those two variables expressed as

$$SROCC = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}, \qquad (1)$$

where d_i denotes the difference between the two ranks of each observation and n is the number of observations.

Another rank order metric is Kendall's τ (tau) measuring the ordinal association between two measured quantities. It can be defined as

$$KROCC = 2 \times \frac{P - N}{n(n-1)}, \qquad (2)$$

where P and N are the numbers of concordant pairs and discordant pairs respectively. All three coefficients calculated for the subjective and objective scores obtained for the same series of images should tend to 1 for a perfect IQA metric. Since the Human Visual System is highly nonlinear, according to the recommendation [31] of the Video Quality Experts Group (VQEG), the nonlinear mapping between the subjective and objective scores, preferably using the logistic function, should be applied.

B. Fusion of IQA Metrics

Since a perfect automatic IQA metric should provide the opportunity to compute the metric's value, which is linearly correlated with subjective perception of various kinds of distortions, the application of any additional nonlinear mapping is not desirable in general. Considering the necessity of choice of some parameters (and type) of the nonlinear mapping function as the result of optimization for a chosen database, the direct application and proper interpretation of the raw metric scores by the user may be troublesome.

As some metrics presented in Section II utilize different types of information, their combination should provide some new possibilities leading to better correlation with subjective scores. Therefore a perfect situation will be the possibly highest correlation of the raw values of the IQA metric with MOS values, especially considering Pearson's correlation, since the rank order correlation coefficients will not change assuming the monotonicity of the mapping function.

The first approach to the use of the combined metrics for this purpose was presented in the paper [32], where the weighted product of three metrics: MS-SSIM, VIF and R-SVD was proposed, leading to the increase of the PCC value from 0.784 (for MS-SSIM as the best from three metrics considering the raw scores) to 0.86 for TID2018 database. Replacing the R-SVD with FSIMc, a further increase of the PCC values to 0.8752 for the same database was achieved for CISI metric [33].

A nonlinear combination of RFSIM and FSIMc metrics [34] led to further increase of the PCC for TID2008 to 0.886 and the application of Extended Hybrid Image Similarity (EHIS) [35], being the optimized combination of four metrics: MS-SSIM, VIF, weighted FSIMc and RFSIM, increased this value to 0.9105. The weighted FSIMc values were obtained optimizing the α and β parameters of (4), as well as the γ parameter for the formula used for I and Q chrominance channels [36].

The verification of this approach for images containing multiple distortions was presented in the paper [37], where the increase from 0.8083 for the best single metric to 0.8596 for the combination of 4 metrics was reported for 270 images from LIVE Multiply Distorted IQA Database. The graphical illustration of the obtained improvement is shown in Fig. 1, presenting the results obtained in the paper [37] for both parts of this dataset (only 270 images out of 450 contain multiple distortions).

Another successful attempt to the multi-metric fusion was presented by M. Oszust in the paper [38], where an optimization problem of a decision fusion was solved using a genetic algorithm for 18 IQA measures, leading to further extension of the approach discussed above, verified also for the newer TID2013 database. Some other results were presented in the paper [39], where more metrics were used, although as the weighted sum instead of the weighted product. Unfortunately, such an approach requires the additional nonlinear mapping with the regression function according to VQEG recommendation [31]. To compare the

results for TID2008 database with previous attempts, the best LCSIM2 metric, consisting of 11 different single IQA metrics, achieves the value of PCC = 0.9202 (after nonlinear regression). Another family of regression based Similarity measures (rSIMs) was presented in [40] with the linear combination of 16 measures leading to PCC = 0.9218 for TID2018 dataset.
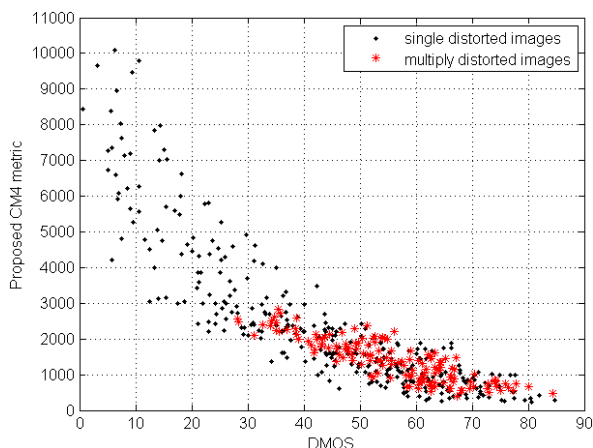

Fig. 1. Scatter plot of the combined metric versus DMOS values for the LIVE Multiply Distorted Image Quality Database [37].

The application of neural networks and some other machine learning algorithms for the fusion of IQA metrics was discussed by Barri *et al.* [41], whereas the use of pairwise score differences to obtain the lasso regression Similarity measures (lrSIMs) was examined by Oszust [42]. Nevertheless, the obtained results were slightly worse than those presented in [40].

Some other approaches to IQA metrics fusion include the application of support vector regression [43], use of neural networks [44], as well as the utilization of SVM classifier for k-nearest-neighbour regression with the fusion of SSIM, VSNR, and VIF [45].

IV. REFERENCELESS (BLIND) METRICS

In many practical applications, particularly related to image and video transmission, the reference images may be unavailable, limiting the applicability of FR metrics. Despite the lower universality and correlation with subjective scores, no-reference metrics seem to be more attractive in such areas of technology [46].

Most of the "blind" metrics are sensitive to only one or two types of distortions, e.g. blur [47]–[50], noise [51], JPEG [52] or JPEG2000 [53] compression artifacts. Nevertheless, some of no-reference metrics, e.g. NIQE, BLIINDS, DIIVINE or BRISQUE [54], utilize also natural scene statistics, anisotropy [55], [56] or entropy [57], [58].

Recently, some other methods were introduced, e.g. based on the use of multiple bandpass and redundancy domains to acquire the complementary features in multiple colour spaces [59], where the Gaussian Mixture Models (GMMs) followed by Fisher Vectors were used to fit them. Another idea [60] was the use of image statistics and robust feature descriptors based on Speeded-Up Robust Features (SURF) approach with a support vector regression technique used for mapping into subjective scores. The additional filtration of the input images was proposed in [61], followed by feature extraction and SVR optimization. In general, many

of such recent metrics are based on machine learning and various image statistics.

Some other representative no-reference IQA metrics include the use of histogram-based features with Laplacian of Gaussian (LoG) responses [62], Oriented Gradients Image Quality Assessment (OG-IQA) [63], histogram of local binary pattern (LBP) based descriptor [64] and other LBP based metrics [65], [66]. An interesting idea of the use of "pseudo-reference image" generated from the distorted image was introduced recently [67]. The Authors developed distortion-specific metrics to estimate blockiness, sharpness, and noisiness, leading to general-purpose blind PRI-based (BPRI) metric. An exemplary application of deep convolutional neural networks for "blind" IQA was presented in the paper [68]. The algorithm proposed by its Authors consists of distortion type classification, CNN based IQA algorithms and fusion algorithm.

One of the most recent ideas is the application of the analysis of distributions of local gradient orientations in image regions of different sizes [69], outperforming even some deep learning based IQA methods.

V. IMAGE QUALITY ASSESSMENT DATABASES

To verify the performance of various metrics, several IQA databases were provided by various groups of researchers. They usually contain several reference images and their distorted versions, subjected to various types and different amount of contaminations, together with results of subjective evaluation by independent human observers, expressed as Mean Opinion Scores (MOS) or Differential MOS (DMOS) values. Knowing the subjective opinions, obtained after statistical analysis and outlier removal, one can calculated the correlation coefficients between the MOS or DMOS values and the currently developed objective metrics. A relevant common problem of the most datasets is that despite they contain colour images the distortions introduced to the subjectively assessed images are not colour specific.

The first popular dataset was the LIVE Image Quality Assessment Database delivered by Laboratory for Image & Video Engineering (LIVE) of Texas University at Austin. Release 2 of this database, with realigned subjective quality data [70], contains 29 original images subjected to five types of distortions: JPEG compression, JPEG2000 compression, white noise, Gaussian blur and simulated fast fading Rayleigh channel (bit errors typical for the wireless transmission of JPEG2000 compressed images) with various intensities or compression ratios. The opinions of human observers were expressed as 982 DMOS values within the range 0–100, however 203 of them regard the original images and therefore they may be often removed from further analysis, since the obtained values of the objective metrics would be evident (e.g. SSIM = 1).

The increasing necessity of a reliable verification of some newer IQA methods caused the development of some other datasets, including the most popular and the largest Tampere Image Datasets. The first one – TID2008 – contains 1700 color images, i.e. 25 reference images subjected to 17 types of distortions with 4 levels each [71]. The MOS values are the results of 838 experiments carried out by observers from three countries: Finland, Italy, and Ukraine with totally

256428 comparisons of visual quality of distorted images.

A newer version of this dataset (TID2013) contains images with 24 types of distortions, including colour specific ones, e.g. change of colour saturation, image colour quantization with dither or chromatic aberrations [72], which can also be analysed in 6 groups. All of the distortions have five levels instead of four used in TID2008. As stated by the Authors, 985 experiments were performed in five countries (Ukraine, Finland, France, USA, and Italy), with further rejection of 14 abnormal experimental results. It is worth noting that some individual distortions, e.g. lossy compression of noisy images, can also be considered as a first step towards the quality assessment of multiply distorted images.

Since the presence of multiple distortions in a single image may cause different subjective opinions, the verification and benchmarking of objective IQA metrics in this regard can be made using LIVE Multiply Distorted Image Quality Database [73]. In fact, images observed by consumers usually reach them after several stages of image processing, which may introduce different distortions, and therefore such judgments are crucial for further development of IQA algorithms. Exemplary images from this database are illustrated in Fig. 2, where the original images are shown on the left. The right images in the top two rows contain the images subjected to blurring and JPEG compression, whereas the right images from the two bottom rows contain blur distortions together with the additive noise.
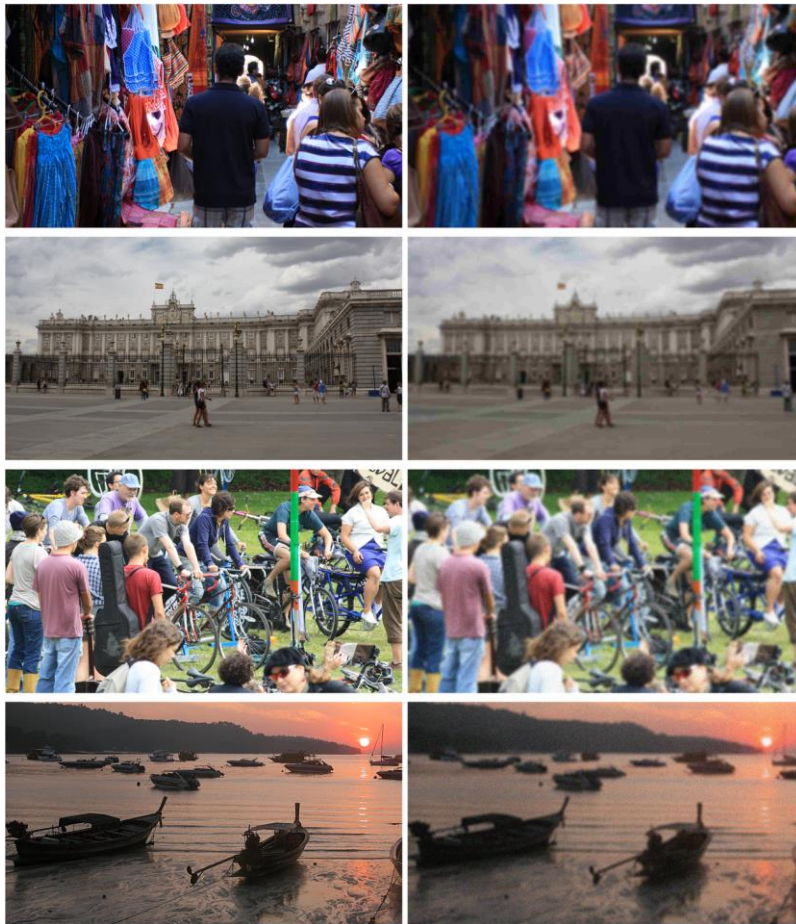

Fig. 2. Exemplary reference (left) and multiply distorted (right) images from LIVE Multiply Distorted Image Quality Database [73].

Some other less popular IQA databases are: Categorical Subjective Image Quality Database (CSIQ) delivered by Oklahoma State University (30 reference images, 6 types of distortions, 866 test images in aggregate, 5000 subjective ratings), IRCCyN/IVC Database from the University of Nantes (10 reference images, 160 distorted images with 4 types of distortions, 15 observers), Wireless Image Quality (WIQ) Database (7 reference greyscale images, 80 test images with distortions typical for wireless transmission, 30 observers), MICT (Toyama) Database (the oldest one with 14 original images and 196 test images corrupted by JPEG and JPEG2000 compression artifacts) and A57/Cornell dataset containing 54 greyscale images obtained from only three reference ones.

A number of dedicated databases, including video, stereoscopic and 3D images, art images, etc. can also be found at the webpages of IRCCyN/IVC and LIVE group. Some recent datasets include: LIVE Mobile Video Quality Database, IDEAL - LIVE Distorted Face Database, LIVE Immersive Image Database, ESPL-LIVE HDR Subjective Image Quality Database or LIVE In the Wild Image Quality Challenge Database.

An interesting original dataset was proposed by Zhai and Neuhoff [74] regarding the similarity of scenic bilevel images, being currently probably the only such database containing binary images. Nevertheless, its importance may be especially high for the development of binary image quality assessment methods, which are rarely analysed, although helpful in many practical applications, including document image analysis and Optical Character Recognition (OCR).

The application of such databases and developed metrics

would supplement some existing methods, which are typically used for evaluation of image binarization results. For this purpose, some dedicated datasets are typically used, containing colour or grayscale images subjected to various distortions together with binary "ground-truth" images. The assessment of results is based more or less on the calculation of the number of properly and improperly classified pixels (true positives, true negatives, false positives and false negatives). They can be further used to determine some typical classification measures, such as e.g. accuracy, specificity, sensitivity, F-Measure, etc. Another possibility, in case of the OCR applications, is their calculation for recognized characters instead of individual pixels, although the proper text data should be known in advance.

Since the Bilevel Image Similarity Ground Truth Archive database is rather less known in image processing community, some exemplary images containing various distortions, influencing the shapes of the visible objects, as well as 7 reference images, are presented in Fig. 3.

Fig. 3. Exemplary images from Bilevel Image Similarity Ground Truth Archive [74]. Two top rows contain the reference images.

Another recently delivered database is the large-scale Waterloo Exploration Database with 4,744 pristine natural images and 94,880 distorted images created from them [75]. Since the collection of individual subjective scores for such large dataset would be very difficult or even impossible, the Authors proposed three alternative test criteria to evaluate the performance of IQA models: pristine/distorted image discriminability test (D-test), the listwise ranking consistency test (L-test), and the pairwise preference consistency test (P-test). All the images in the database were divided into seven major groups: human, animal, plant, landscape, cityscape, still-life and transportation.

VI. Alternative Applications of IQA Methods

Image quality assessment methods, particularly full-reference metrics which can be considered as similarity measures, may be useful for many different purposes. One of such "natural" areas of their applications may be the image and video quality assessment for computer games and computer graphics, as well as computer-generated images [76] with future possible applications in virtual and augmented reality.

In computer games the localization of distortions may be as much important as the overall evaluation, however some of existing metrics can be efficiently applied for this purpose [77]. Therefore, some datasets and metrics with marked and subjectively assessed local distortions, especially game artifacts, are of particular interest [78].

Another recent application of full-reference IQA methods is their use for the evaluation of 3D printed surfaces. Although their direct application for such purposes is limited due to the unavailability of reference 3D prints, as well as the necessity of phase adjustments and colour calibration, the mutual comparison of image fragments is possible. The application of SSIM, CW-SSIM and STSIM metrics with the division of the image of the 3D printed surface into 4 and 16 parts was presented in the paper [79], whereas the application of feature similarity metrics (RFSIM and FSIM) was roughly analysed in [80]. An improved version of the application of SSIM, with the use of the Monte Carlo method to decrease the computational complexity, was proposed in [81], although its efficient application requires the additional phase adjustment of randomly chosen image fragments.

The application of combined metrics for the evaluation of the 3D printed surfaces is one of the obvious directions of further research, possibly in combination with some other investigated approaches.

VII. Summary

Regardless of numerous methods and approaches to image and video quality assessment, this task still remains one of the open challenges in image analysis.

Considering recent successful applications of machine learning approaches, and particularly deep convolutional neural networks (CNNs), in various computer vision and pattern recognition problems, one of the future trends seems to be their application in referenceless image quality assessment. Nevertheless, a significant limitation may be the necessity of data augmentation for training the network, which can be partially solved using the large-scale datasets, such as Waterloo Exploration Database [75].

Another direction of research may be the development of colour IQA methods, conditioned by the wider availability of databases containing colour specific distortions, such as TID2013 [72]. A similar increase of interest of researchers may be expected for multiply distorted images, as well as video quality assessment utilizing inter-frame similarity instead of the classical frame-by-frame approaches.

New image similarity metrics should also be stimulating for the development of texture similarity methods and better CBIR algorithms. They can also be partially useful for machine vision supported navigation and self-localization of mobile robots and autonomous vehicles in partially known environment. Texture analysis methods, also utilizing texture similarity assessment, may also be useful for terrain classification for autonomous landing of drones [82].

REFERENCES

[1] Z. Wang, A.C. Bovik, "A universal image quality index", *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002. DOI: 10.1109/97.995823.

[2] Z. Wang, A. C. Bovik, H. Sheikh, E. Simoncelli, "Image quality assessment: From error measurement to structural similarity", *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.

[3] Z. Wang, E. Simoncelli, A. C. Bovik, "Multi-scale structural similarity for image quality assessment", in *Proc. 37th IEEE Asilomar Conf. Signals, Systems and Computers*, vol. 2, Pacific Grove, California, 2003, pp. 1398–1402. DOI: 10.1109/ACSSC.2003. 1292216

[4] G.-H. Chen, C.-L. Yang, S.-L. Xie, "Gradient-based Structural Similarity for image quality assessment", in *Proc. IEEE Int. Conf. Image Processing (ICIP 2006)*, Altanta, Georgia, 2006, pp. 2929–2932. DOI: 10.1109/ICIP.2006.313132.

[5] S. Aja-Fernandez, R. San-Jose-Estepar, C. Alberola-Lopez, C. Westin, "Image quality assesment based on local variance", in *Proc. 28th Annu. IEEE Int. Conf. Engineering in Medicine and Biology Society (EMBS 2006)*, New York, 2006, pp. 4815–4818. DOI: 10.1109/IEMBS.2006.259516.

[6] W. Lin, M. Narwaria, "Image quality assessment based on gradient similarity", *IEEE Trans. Image Processing*, vol. 21 no. 4, pp. 1500–1512, 2012. DOI: 10.1109/TIP.2011.2175935.

[7] C. Li, A.C. Bovik, "Three-component weighted structural similarity index", in *Proc. SPIE,* vol. 7242 – *Image Quality and System Performance VI*, 72420Q, 2009. DOI: 10.1117/12.811821.

[8] M.-J. Chen, A.C. Bovik, "Fast structural similarity index algorithm", *J. Real-Time Image Processing*, vol. 6 no. 4, pp. 1–7, 2011. DOI: 10.1007/s11554-010-0170-9.

[9] D. Brunet, E. R. Vrscay, Z. Wang, "On the mathematical properties of the Structural Similarity index", *IEEE Trans. Image Processing*, vol. 21 no. 4, pp. 1488–1499, 2012. DOI: 10.1109/TIP.2011.2173206.

[10] A. Rehman, Z. Wang, "Reduced-reference image quality assessment by Structural Similarity estimation", *IEEE Trans. Image Processing*, vol. 21 no. 8, pp. 3378–3389, 2012. DOI: 10.1109/TIP.2012.2197011.

[11] A. Kolaman, O. Yadid-Pecht, "Quaternion Structural Similarity: a new quality index for color images", *IEEE Trans. Image Processing*, vol. 21 no. 4, pp. 1526–1536, 2012. DOI: 10.1109/TIP.2011.2181522.

[12] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, M. K. Markey, "Complex wavelet structural similarity: a new image similarity index", *IEEE Trans. Image Processing*, vol. 18 no. 11, pp. 2385–2401, 2009. DOI: 10.1109/TIP.2009.2025923.

[13] Z. Wang, Q. Li, "Information content weighting for perceptual image quality assessment", *IEEE Trans. Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011. DOI: 10.1109/TIP.2010.2092435.

[14] X. Zhang, X. Feng, W. Wang, W. Xue, "Edge strength similarity for image quality assessment", *IEEE Signal Proc. Letters*, vol. 20 no. 4, pp. 319–322, 2013, DOI: 10.1109/LSP.2013.2244081.

[15] L. Zhang, L. Zhang, X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms", in *Proc. 17th IEEE Int. Conf. on Image Processing (ICIP 2010)*, Hong Kong, China, 2010, pp. 321–324. DOI: 10.1109/ICIP.2010.5649275.

[16] L. Zhang, L. Zhang, X. Mou, D. Zhang, "FSIM: A feature similarity index for image quality assessment", *IEEE Trans. Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011. DOI: 10.1109/TIP.2011.2109730.

[17] J. Zujovic, T. N. Pappas, D. L. Neuhoff, "Structural texture similarity metrics for image analysis and retrieval", *IEEE Trans. Image Processing*, vol. 22, no. 7, pp. 2545–2558, 2013. DOI: 10.1109/TIP.2013.2251645.

[18] H. Sheikh, A. C. Bovik, G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics", *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005. DOI: 10.1109/TIP.2005.859389.

[19] H. Sheikh, A. C. Bovik, "Image information and visual quality", *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 430–444, 2006. DOI: 10.1109/TIP.2005.859378.

[20] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, A. C. Bovik, "Image quality assessment based on a degradation model", *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 636–650, 2000. DOI: 10.1109/83.841940.

[21] D. M. Chandler, S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images", *IEEE Trans. Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007. DOI: 10.1109/TIP.2007.901820.

[22] R. Reisenhofer, S. Bosse, G. Kutyniok, T. Wiegand, "A Haar wavelet-based perceptual similarity index for image quality assessment", *Signal Processing: Image Communication*, vol. 61, pp. 33–43, 2018. DOI: 10.1016/j.image.2017.11.001.

[23] W. Sun, Q. Liao, J. H. Xue, F. Zhou, "SPSIM: a SuperPixel-based SIMilarity index for full-reference image quality assessment", *IEEE Trans. Image Processing*, vol. 27, no. 9, pp. 4232–4244, 2018. DOI: 10.1109/TIP.2018.2837341.

[24] A. Shnayderman, A. Gusev, A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment", *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 422–429, 2006. DOI: 10.1109/TIP.2005.860605.

[25] A. Mahmoudi-Aznaveh, A. Mansouri, F. Torkamani-Azar, M. Eslami, "Image quality measurement besides distortion type classifying", *Optical Review*, vol. 16, no. 1, pp. 30–34, 2009. DOI: 10.1007/s10043-009-0007-6.

[26] A. Mansouri, A. Mahmoudi-Aznaveh, F. Torkamani-Azar, J. A. Jahanshahi, "Image quality assessment using the singular value decomposition theorem", *Optical Review*, vol. 16, no. 2, pp. 49–53, 2009. DOI: 10.1007/s10043-009-0010-y.

[27] S. Wang, C. Deng, W. Lin, B. Zhao, J. Chen, "A novel SVD-based image quality assessment metric", in *Proc. 2013 IEEE Int. Conf. Image Processing (ICIP 2013)*, Melbourne, Australia, pp. 423–426, 2013. DOI: 10.1109/ICIP.2013.6738087.

[28] A. Hu, R. Zhang, D. Yin, Y. Zhan, "Image quality assessment using a SVD-based structural projection", *Signal Processing: Image Communication*, vol. 29, no. 3, pp. 293–302, 2014. DOI: 10.1016/j.image.2014.01.007.

[29] A. Mansouri, A. Mahmoudi-Aznaveh, "SSVD: structural SVD-based image quality assessment", *Signal Processing: Image Communication*, vol. 16, no. 2, pp. 49–53, 2019. DOI: 10.1016/j.image.2019.01.007.

[30] M. Gaubatz, "MeTriX MuX visual quality assessment package". Online. [Available]: http://ollie-imac.cs.northwestern.edu/~ollie/GMM/ code/metrix_mux/

[31] VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II)", Technical Report, Video Quality Expert Group, 2003. Online. [Available]: https://www.itu.int/md/T01-SG09-C-0060

[32] K. Okarma, "Combined full-reference image quality metric linearly correlated with subjective assessment", *Lecture Notes in Computer Science*, vol. 6113, 2010, pp. 539–546. DOI: 10.1007/978-3-642-13208-7_67.

[33] K. Okarma, "Combined image similarity index", *Optical Review*, vol. 19, no. 5, pp. 249–254, 2012. DOI: 10.1007/s10043-012-0055-1.

[34] K. Okarma, "Hybrid feature similarity approach to full-reference image quality assessment", *Computer Vision and Graphics*, *Lecture Notes in Computer Science*, pp. 212–219, 2012. DOI: 10.1007/ 978-3-642-33564-8_26.

[35] K. Okarma, "Extended hybrid image similarity – combined full-reference image quality metric linearly correlated with subjective scores", *Elektronika ir Elektrotechnika*, vol. 19, no. 10, pp. 129–132, 2013. DOI: 10.5755/j01.eee.19.10.5908.

[36] K. Okarma, "Weighted Feature Similarity – a nonlinear combination of gradient and phase congruency for full-reference image quality assessment", *Image Processing and Communications Challenges 4*, vol. 184, pp. 187–194, 2013. DOI: 10.1007/978-3-642-32384-3_23.

[37] K. Okarma, "Quality assessment of images with multiple distortions using combined metrics", *Elektronika ir Elektrotechnika*, vol. 20, no. 6, pp. 128–131, 2014. DOI: 10.5755/j01.eee.20.6.7284.

[38] M. Oszust, "Decision fusion for image quality assessment using an optimization approach", *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 65–69, 2016, DOI: 10.1109/LSP.2015.2500819.

[39] M. Oszust, "Full-reference image quality assessment with linear combination of genetically selected quality measures", *PLoS ONE*, vol. 11, no. 6, 2016. DOI: 10.1371/journal.pone.0158333.

[40] M. Oszust, "A regression-based family of measures for full-reference image quality assessment", *Measurement Science Review*, vol. 16, no. 6, pp. 316–325, 2016. DOI: 10.1515/msr-2016-0040.

[41] A. Barri, A. Dooms, B. Jansen, P. Schelkens, "A locally adaptive system for the fusion of objective quality measures", *IEEE Trans. Image Processing*, vol. 23, no. 6, pp. 2446–2458, 2014. DOI: 10.1109/TIP.2014.2316379.

[42] M. Oszust, "Image quality assessment with lasso regression and pairwise score differences", *Multimedia Tools and Applications*, vol. 76, no. 11, pp. 13255–13270, 2017. DOI: 10.1007/s11042-016-3755-x.

[43] T-J. Liu, W. Lin, C-C. J. Kuo, "Image quality assessment using multi-

method fusion", *IEEE Trans. Image Processing*, vol. 22, no. 5, pp. 1793–1807, 2013. DOI: 10.1109/TIP.2012.2236343.

[44] V. V. Lukin, N. N. Ponomarenko, O. I. Ieremeiev, K. O. Egiazarian, J. Astola, "Combining full-reference image visual quality metrics by neural network", *SPIE/IS&T Electronic Imaging*, San Francisco, California, United States, 2015, p. 93940K. DOI: 10.1117/12.2085465.

[45] P. Peng, Z. N. Li, "A mixture of experts approach to multi-strategy image quality assessment", *Image Analysis and Recognition, Lecture Notes in Computer Science*, vol. 7324, pp. 123–130, 2012. DOI: 10.1007/978-3-642-31295-3_15.

[46] X. Li, "Blind image quality assessment", in *Proc. IEEE Int. Conf. Image Processing (ICIP 2002)*, Rochester, New York, 2002, pp. 449–452. DOI: 10.1109/ICIP.2002.1038057.

[47] J. Caviedes, S. Gurbuz, "No-reference sharpness metric based onlocal edge kurtosis", in *Proc. IEEE Int. Conf. Image Processing (ICIP 2002)*, Rochester, New York, 2002, vol. 3, pp. 53–56. DOI: 10.1109/ICIP.2002.1038901.

[48] S. Varadarajan, L. J. Karam, "An improved perception-based no-reference objective image sharpness metric using iterative edge refinement", in *Proc. 15th IEEE Int. Conf. Image Processing (ICIP 2008)*, San Diego, CA, USA, 2008, pp. 401–404. DOI: 10.1109/ICIP.2008.4711776.

[49] N. D. Narvekar, L. J. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection", in *Proc. 1st IEEE Int. Workshop on Quality of Multimedia Experience (QoMEX 2009)*, San Diego, CA, USA, 2009, pp. 87–91. DOI: 10.1109/QOMEX.2009.5246972.

[50] R. Ferzli, L. J. Karam, "A no-reference objective image sharpness metric based on the notion of Just Noticeable Blur (JNB)", *IEEE Trans. Image Processing*, vol. 18, no. 4, pp. 717–728, 2009. DOI: 10.1109/TIP.2008.2011760.

[51] T. Zhu, L. J. Karam, "A no-reference objective image quality metric based on perceptually weighted local noise", *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, 2014. DOI: 10.1186/1687-5281-2014-5.

[52] Z. Wang. A. C. Bovik, B. L. Evan, "Blind measurement of blocking artifacts in images", in *Proc. 2000 IEEE Int. Conf. Image Processing (ICIP 2000)*, Vancouver, Canada, vol. 3, pp. 981–984, 2000. DOI: 10.1109/ICIP.2000.899622.

[53] Z. M. Parvez Sazzad, Y. Kawayoke, Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features", *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 257–268, 2008. DOI: 10.1016/j.image.2008.03.005.

[54] A. Mittal, A. K. Moorthy, A. C. Bovik, "No-reference image quality assessment in the spatial domain", *IEEE Trans. Image Processing*, vol. 21, no. 12, pp. 4659–4708, 2012. DOI: 10.1109/TIP.2012.2214050.

[55] S. Gabarda, G. Cristobal, "Blind image quality assessment through anisotropy", *J. Opt. Soc. Am. A*, vol. 24, no. 12, pp. B42–B51, 2007. DOI: 10.1364/JOSAA.24.000B42.

[56] S. Gabarda, G. Cristobal, N. Goel, "Anisotropic blind image quality assessment: Survey and analysis with current methods", *J. Vis, Comm. Image Representation*, vol. 52, pp. 101–105, 2018. DOI: 10.1016/j.jvcir.2018.02.008.

[57] S. Gabarda, R. Redondo, E. Gil, G. Cristobal, "Image denoising and quality assessment through the Renyi entropy", in *Proc. SPIE, vol. 7444 – Mathematics for Signal and Information Processing*, 2009, p. 744419. DOI: 10.1117/12.826153.

[58] X. Yang, F. Li, W. Zhang, L. He, "Blind image quality assessment of natural scenes based on entropy differences in the DCT domain", *Entropy*, vol. 20, no. 11, 2018. DOI: 10.3390/e20110885.

[59] Y. Ma, W. Zhang, J. Yan, C. Fan, W. Shi, "Blind image quality assessment in multiple bandpass and redundancy domains", *Digital Signal Processing*, vol. 80, pp. 37–47, 2018. DOI: 10.1016/j.dsp.2018.05.010.

[60] M. Oszust, "No-Reference image quality assessment using image statistics and robust feature descriptors", *IEEE Signal Processing Letters*, vol. 21, no. 11, pp. 1656–1660, 2017. DOI: 10.1109/LSP.2017.2754539.

[61] M. Oszust, "Optimized filtering with binary descriptor for blind image quality assessment", *IEEE Access*, vol. 6, pp. 42917–42929, 2018. DOI: 10.1109/ACCESS.2018.2860127.

[62] W. Xue, X. Mou, L. Zhang, A. C. Bovik, X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features", *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, 2014. DOI: 10.1109/TIP.2014.2355716.

[63] L. Liu, Y. Hua, Q. Zhao, H. Huang, A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network", *Signal Processing: Image Communication*, vol. 40, pp. 1–15, 2016. DOI: 10.1016/j.image.2015.10.005.

[64] Q. Li, W. Lin, Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain", *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, 2016. DOI: 10.1109/LSP.2016.2537321.

[65] Q. Li, W. Lin, Y. Fang, "BSD: Blind image quality assessment based on structural degradation", *Neurocomputing*, vol. 236, pp. 93–103, 2017. DOI: 10.1016/j.neucom.2016.09.105.

[66] Q. Li, W. Lin, J. Xu, Y. Fang, "Blind image quality assessment using statistical structural and luminance features", *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, 2016. DOI: 10.1109/TMM.2016.2601028.

[67] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, C. W. Chen, "Blind quality assessment based on pseudo-reference image", *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2049–2062, 2018. DOI: 10.1109/TMM.2017.2788206.

[68] C. Fan, Y. Zhang, L. Feng, Q. Jiang, "No reference image quality assessment based on multi-expert convolutional neural networks", *IEEE Access*, vol. 6, pp. 8934–8943, 2018. DOI: 10.1109/ACCESS.2018.2802498.

[69] M. Oszust, "No-reference image quality assessment with local gradient orientations", *Symmetry*, vol. 11, no. 1, 2019. DOI: 10.3390/sym11010095.

[70] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, "TID2008 – a database for evaluation of full-reference visual quality assessment metrics", *Advances of Modern Radioelectronics*, vol. 10, pp. 3045, 2009.

[71] 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, pp. 57–77, 2015. DOI: 10.1016/j.image.2014.10.009.

[72] D. Jayaraman, A. Mittal, A. K. Moorthy, A. C. Bovik, "Objective image quality assessment of multiply distorted images", in *Conf. Rec. 46th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, 2012, pp. 1693–1697. DOI: 10.1109/ACSSC.2012.6489321.

[73] Y. Zhai, D. Neuhoff, "Similarity of scenic bilevel images", *IEEE Trans. Image Processing*, vol. 25, no. 11, pp. 5063–5076, 2016. DOI: 10.1109/TIP.2016.2598493.

[74] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, L. Zhang, "Waterloo Exploration Database: new challenges for image quality assessment models", *IEEE Trans. Image Processing*, vol. 26, no. 2, pp. 1004–1016, 2017. DOI: 10.1109/TIP.2016.2631888.

[75] A. Bigand, J. Dehos, C. Renaud, J. Constantin, "Image quality assessment of computer-generated images based on machine learning and soft computing", *Springer Briefs in Computer Science*, 2018. DOI: 10.1007/978-3-319-73543-6.

[76] R. Piorkowski, R. Mantiuk, A. Siekawa, "Automatic detection of game engine artifacts using full reference image quality metrics", *ACM Trans. Applied Perception*, vol. 14, no. 3, 2017. DOI: 10.1145/3047407.

[77] K. Wolski, D. Giunchi, N. Ye, P. Didyk, K. Myszkowski, R. Mantiuk, H.-P. Seidel, A. Steed, R. K. Mantiuk, "Dataset and metrics for predicting local visible differences", *ACM Trans. Graphics*, vol. 37, no. 5, 2018. DOI: 10.1145/3196493.

[78] K. Okarma, J. Fastowicz, M. Teclaw, "Application of structural similarity based metrics for quality assessment of 3D prints", in *Computer Vision and Graphics, Lecture Notes in Computer Science*, vol. 9972, pp. 244–252, 2016. DOI: 10.1007/978-3-319-46418-3_22.

[79] K. Okarma, J. Fastowicz, "Quality assessment of 3D prints based on feature similarity metrics", in *Image Processing and Communications Challenges 8. Advances in Intelligent Systems and Computing*, vol. 525, pp. 104–111, 2017. DOI: 10.1007/978-3-319-47274-4_12.

[80] J. Fastowicz, K. Okarma, "Fast quality assessment of 3D printed surfaces based on structural similarity of image regions", in *Proc. 2018 Int. Interdisciplinary PhD Workshop*, Swinoujscie, Poland, 2018, pp. 401–406. DOI: 10.1109/IIPHDW.2018.8388399.

[81] M. Skoczylas, "Vision analysis system for autonomous landing of micro drone", *Acta Mechanica et Automatica*, vol. 8, no. 4, pp. 199–203, 2014. DOI: 10.2478/ama-2014-0036.