Improving the criteria for quality assessment of image processing algorithms

D V Dubinin [1], A I Kochegurov [2] and V E Geringer [3*]
[1]Tomsk State University of Control Systems and Radioelectronics, Lenin Avenue 40, Tomsk, 634050, Russian Federation
[2]National Research Tomsk Polytechnic University, Lenin Avenue 30, Tomsk, 634050, Russian Federation
[3]Baden-Wuerttemberg Cooperative State University, Faculty of Engineering, Fallenbrunnen 2, D-88045, Friedrichshafen, Germany

* geringer@dhbw-ravensburg.de

Abstract. Quality criteria are the essence of a measurement system. The goal of assessing the quality of software elements (algorithms) used for image processing is to ensure control of technical performance indicators of the system as a whole or its individual functional units under the reduction of costs associated with minimizing the loss function (tuning and debugging). The correct choice of individual metrics for the generalized quality indicator to solve tasks in a particular subject area is one of the key steps in system optimization. It ensures the most flexible approach to testing the developed software elements, identifying and eliminating their functional shortcomings. At the same time, we can say that in this way the generalized indicator of the entire system (its goal function) is optimized.

1. Introduction
Nowadays, improvement in the quality of aviation and space technology is often associated with the creation of highly efficient hardware and software systems for survey and measurement, monitoring of the Earth's surface, geoinformation systems and avionics. The software systems are often operated in abnormal conditions that are dependent on light, sensors and recording systems, signal-to-noise ratio, noise statistics, sampling intervals, which imposes certain technical and quality requirements at all stages [1]. This makes quality assessment a rather complex task [2 -5], which, once failed, often leads to significant financial losses and failures at certain stages of design and implementation. Refinement of a unified sequence of quality assessment and monitoring into a clear system of testing and eliminating shortcomings of individual functional blocks is a fundamental task at all design and implementation stages. This approach allows achieving the set goal, creating an efficient and easy-to-use system that ensures the uniformity and reliability of measurements and performance monitoring.

2. Objectives and research methods
To carry out a qualitative and quantitative assessment of a system, it is necessary to determine a number of basic indicators that are fundamental in terms of achieving basic technical characteristics. The purpose of this work is to show possible approaches and principles of creating an efficient unified system

for assessing the quality of image processing, which ensures the uniformity and reliability of measurements and can be used to design improved quality criteria.

The main research methods include empirical and theoretical studies, such as programming, statistical approaches, experimental design, simulation, data visualization (functions, graphs, etc.), detection and localization of contour drawing of images, formal representations of systems, etc. A special approach was based on the idea of gradual formalization of models by alternately using means and methods of activating the intuition of specialists, as well as methods of collecting, storing and processing information.

Computer modeling was carried out using the Windows 10 operating system in the Delphi environment; the numerical models were validated in the Mathcad Software environment.

3. Formalization of the task

Let us consider task formalization by the example of developing a quality criterion for contour drawing of an image [6, 7]. Summarizing the existing methods to assess the quality of image processing, two main approaches can be distinguished: 1) objective quality assessment of image processing. 2) subjective quality assessment based on expert assessments by various methods, e.g. Mean Opinion Score (MOS).

Both approaches have a number of advantages and disadvantages. We will focus on objective quality assessment of image processing where calculations should be carried out in real time and the system should cope with a sequence of images arriving from hardware.

Reference images were prepared in a KIM SP software package environment [8] on the basis of a two-dimensional recurrent point renewal stream. Reference images had different morphological structures. Statistical data were generated based on morphologies of types A, F and FD.

The use of the two-dimensional recurrent point renewal stream guaranteed a continuous contour drawing of an image and made it possible to analyze the detection quality of individual components of some selected boundary [9], whereat the image boundary was recognized at a sharp change in brightness.

The properties of the generated reference images were set by the a-priory operator at the stage of setting up and carrying out the experiment [8, 9]. The generation of reference images using a recurrent point renewal stream and the formation of statistics of artificial reference images made it possible to sett up an experiment according to Fisher principle (the principle of randomization). By varying model parameters, an additive noise component with a normal distribution law was introduced. The mathematical expectation of the noise was zero, and the signal-to-noise ratio varied from 1 to 48 dB.

The processing of the reference images by different drawing operators made it possible to obtain an a-posteriori vector description of the passage of the contour of a certain morphological structure of the image [10, 11]. Analysis of the main errors by operators of boundary detection and localization, as well as certain requirements from the point of view of 'functional efficiency' made it possible to make the correct choice of individual metrics of the generalized quality indicator and contributed to the control of technical indicators of the system operability [12]. Comparison of the a-priory vector description with the vector description obtained during the experiment lays foundation for research, development and improvement of individual objective metrics. The study of the generalized quality indicators that include several metrics (e.g. functional efficiency, etc.) and a binary classification present no major problem. Establishing requirements for a specific technical level (requirements analysis and design, model development, carrying out experiment, etc.) and the quality of the final stage of the experiment (checking the correspondence between the numerical model and its mathematical formalization, developing statistics, principles for assessment and analysis of results, etc.) allows developing a unified sequence that forms a clear testing system, is easy to use and ensures the consistency and reliability of quality measurements.

4. Results and recommendations

Analyzing the well-known Pratt's Figure of Merit [13], it is easy to make sure that peak values are formed at low signal-to-noise ratios (PSNR: Peak Signal-to-Noise Ratio, see Figure 1 (right)).

Yaskorsky [14] demonstrated the possibility and the ways to improve the quality indicator proposed by Pratt. From our point of view, the count of errors due to local bias proposed by Pratt (error type: LE – Localization error) and broken boundaries proposed by Yaskorsky (error type: OE – Ommision error) do not fully represent all types of errors.

Therefore, we proposed to introduce another component that is similar to the Pratt's Figure of Merit and takes into account the blurring or thickening of the boundaries (error type: MRE – multiples responses error). Thus, all the main errors were taken into account.

Generalization of the three metrics was carried out according to the multiplicative principle and was presented in the form [6]:

$$MPFoM = FoM_{LE} \cdot FoM_{OE} \cdot FoM_{MRE} \tag{1}$$

where *MPFoM* is a modified Pratt's Figure of Merit [4], see Fig. 1,a.

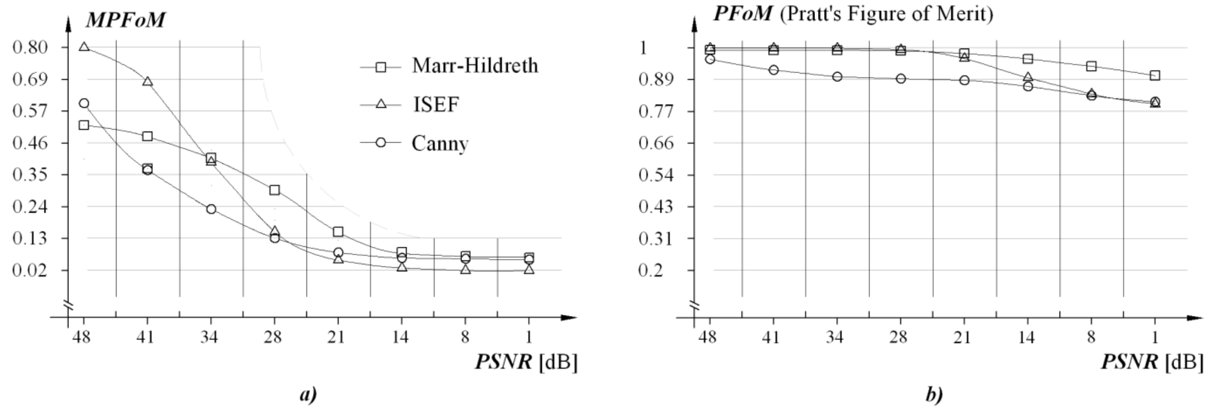

Figure 1. Quality assessment of the contour drawing by a modified (a) and conventional (b) Pratt's Figure of Merit at different signal-to-noise ratios.

Figure 1 shows the results of assessing the quality of the contour drawing by three detectors (Canny, Marr-Hildreth and ISEF) for the two criteria under consideration. It can be seen that the modified Pratt's Figure of Merit provides the most objective quality assessment of the detector performance. Then, the probabilities of type I errors (see Fig. 2, a) and type II errors (see Fig. 2, b) were estimated for the detectors based on binary classification [15].

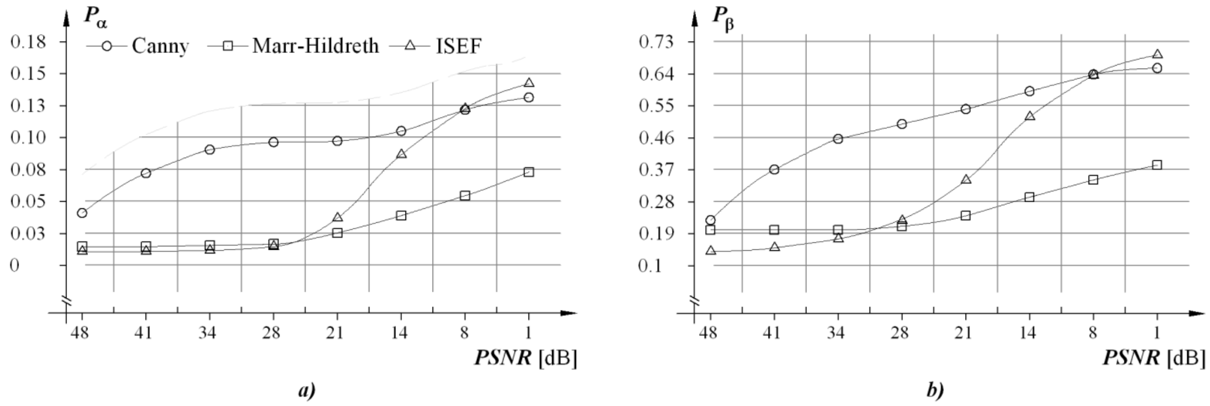

Figure 2. Probabilities of type I errors and type II errors for three operators of boundary detection and localization at different signal-to-noise ratios.

Figure 3 shows the averaged estimates of the power function for three operators of boundary detection and localization at different signal-to-noise ratios.

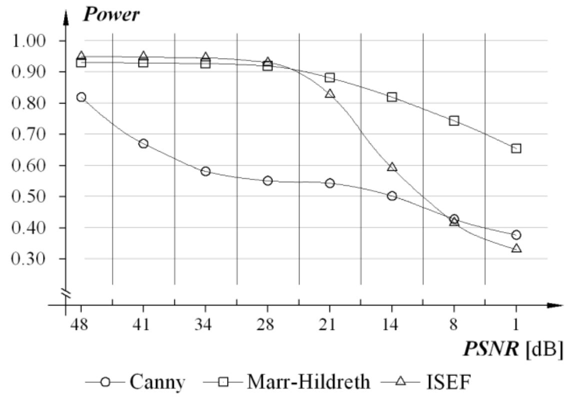

Figure 3. Estimates of the power function for three operators of boundary detection and localization at different signal-to-noise ratios.

The results of calculating detection errors shown in Figures 2 and 3 indicate the correct use of the operators of boundary detection and localization in the modified Pratt's Figure of Merit.

Therefore, the modified Pratt's Figure of Merit can be considered as a successfully improved criterion for quality assessment of image detection algorithms. At the same time, the individual components of the generalized quality indicator should not be mutually dependent, and the quality indicator itself should not have a large number of metrics, e.g. no more than 4–5 metrics. It is also necessary to use a qualitative approach when choosing weight coefficients in the case of the additive formation of a generalized quality indicator. To ensure the reliability of quality measurements, it is necessary to conduct a complex assessment – monitor the final stage of the experiment, e.g. based on discriminatory power, averaged Youden's index, type II errors, etc. depending on the signal-to-noise ratios.

5. Conclusion

A step-by-step integration of generalized quantitative quality measurements includes a sequence of processes that are necessary to achieve some set goals and objectives: to collect and analyze information about the modeling object, choose the modeling environment and determine the properties of individual components of the functional model, carry out mathematical formalization of the model (set up the environmental model), create a numerical model of the experiment and verify it, test the model (for general validation of the modeling system), develop statistics of artificial reference images with certain morphological and statistical properties, set up and conduct an experiment with subsequent analysis of the results. Artificial reference images are used in the experiment to ensure the statistical reliability of the results and the possibility of factor analysis.

The correct choice of the components (metrics) of the generalized quality indicator provides for a detailed analysis of the process of detecting and localizing boundaries, allows the most flexible approach to testing the developed software and hardware systems, identifying and eliminating their shortcomings, and also enables identifying the most reasonable quasi-optimal solutions of different tasks and different functional solutions.

References

[1] Heldman K 2018 Project management: Professional. 5th ed. (Sybex)
[2] Gudkin L 1975 Optimization of electronic devices in terms of a set of quality indicators (M.: Sovetskoe radio) [in Russian]
[3] Magnier B, Abdulrahman H and Montesinos P 2018 A Review of Supervised Edge Detection Evaluation Methods and an Objective Comparison of Filtering Gradient Computations Using Hysteresis Thresholds. *J. Imaging*. **4** 2-63

[4] Dey N, Ashour A, Shi F and Balas V E 2018 *Soft Computing Based Medical Image Analysis* (San Diego: Elsevier Science)

[5] Magnier B 2019 Edge Detection Evaluation: A New Normalized Figure of Merit. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Brighton, United Kingdom.* pp 2407-11

[6] Geringer V, Dubinin D and Kochegurov A 2016 The Results of a Complex Analysis of the Modified Pratt-Yaskorskiy Performance Metrics Based on the Two-Dimensional Markov-Renewal-Process. *Lecture Notes in Computer Science* **9875** 187-196.

[7] Kochegurov A, Dubinin D, Geringer V and Reif K 2019 Estimation of the Efficiency of Contour Detectors on the Basis of a Point Recovery Flow. *Bulletin of the Tomsk Polytechnic University*. **330** 204-216

[8] Dubinin D, Geringer V, Kochegurov A and Reif K 2013 *Bundled software for simulation modeling* (Iasi: ISSCS Press)

[9] Dubinin D, Kochegurov A and Geringer V 2012 On the Statistics of Space-Time Signals Created by a Two-Dimensional Markov Renewal Process. *Bulletin of the Tomsk Polytechnic University* **321** 194-198.

[10] Dubinin D, Geringer V and Kochegurov A 2012 Eine Methode zur Erzeugung stochastischer Helligkeitsfelder durch homogene, einstufige Markoff-Ketten. tm - Technisches Messen. **79** 271-276.

[11] Dmitry D, Geringer V, Kochegurov A and Reif K 2014 Ein stochastischer Algorithmus zur Bildgenerierung durch einen zweidimensionalen Markoff-Erneuerungsprozess. *Automatisierungstechnik* **62** 57-64

[12] Dubinin D, Geringer V, Kochegurov A and Reif K 2014 An efficient method to evaluate the performance of edge detection techniques by a two-dimensional Semi-Markov model. (CICA: IEEE Press)

[13] Abdou I and Pratt W 1979 Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE* **67** 753-763

[14] Yaskorsky A 1987 Criteria for the performance evaluation of the edge detectors. *Avtometria* **3** 127-128 [in Russian]

[15] Starovoitov V and Golub Y 2020 Comparative study of quality estimation of binary classification. *Informatics* **19** 87-101