Ministry of Education of the Republic of Belarus
Educational Institution
Belarusian State University
of Informatics and Radioelectronics

UDC 004. 932

Wang Jing

PROGRAM AND MATHEMETICAL TOOLS FOR RESTORING IMAGES WITH MISSING DATA

ABSTRACT
for a master's degree

Speciality 7-06-0611-02 Information security

Academic Supervisor
German Yuliya Olegovna
Candidate of Technical Sciences,
Associate Professor

Minsk 2026

INTRODUCTION

In the contemporary digital era, images have become an indispensable medium for information exchange, serving critical roles in fields ranging from medical diagnostics and remote sensing to social media communication and forensic investigation. However, the integrity of digital images is frequently compromised during acquisition, transmission, and storage processes, leading to various forms of degradation including noise corruption, blur, and missing data regions. The restoration of images with missing or corrupted data represents a fundamental challenge in digital image processing.

The dissertation discusses distorted data in general, while missing data is a specific case of distorted data. By missing data, we mean the absence of data or the filling of a picture with corrupted data.

The problem of image restoration with missing data is particularly pertinent in real-world scenarios where complete information loss occurs in certain image regions. Unlike noise corruption, which perturbs pixel values while preserving the underlying image structure, missing data scenarios involve the complete absence of information in affected regions. This occurs in numerous practical contexts: medical imaging where artifacts obscure diagnostic regions, satellite imagery where transmission errors cause data loss, historical photograph restoration where physical damage has destroyed image content, and video processing where frame corruption results in missing visual information.

The practical significance of developing robust mathematical tools for image restoration extends across multiple domains. In medical imaging, accurate restoration of corrupted regions can enhance diagnostic capabilities and improve patient outcomes. In surveillance and security applications, the ability to recover missing facial features or license plate information from degraded footage can be crucial for identification purposes. In cultural heritage preservation, restoration techniques enable the recovery of valuable historical visual records that would otherwise be lost to deterioration. Furthermore, in scientific imaging applications, such as astronomical observations and microscopy, restoration methods can recover meaningful data from observations compromised by equipment limitations or environmental interference.

The novelty of this research lies in its systematic approach to comparing and integrating multiple mathematical methodologies for image quality assessment and restoration evaluation. While existing literature has extensively

explored individual restoration techniques, there remains a significant gap in comprehensive comparative analysis that evaluates statistical hypothesis testing methods across diverse distortion types and sample sizes. This research addresses this gap by developing a unified framework that combines the interpretability and rigor of statistical methods.

The relevance of this work is underscored by the increasing demand for automated quality assessment tools in image processing pipelines. As digital image volumes continue to grow exponentially, manual evaluation of restoration quality becomes impractical, necessitating reliable automated methods. The development of mathematical tools that can quantitatively assess whether a restored image maintains statistical equivalence to the original provides a foundation for quality control in automated restoration systems.

This thesis presents a comprehensive investigation into program and mathematical tools for restoring images with distorted data. The research encompasses the theoretical foundations of various distortion models, the implementation and comparison of statistical and machine learning-based assessment methods, and the development of practical software tools for image quality evaluation.

GENERAL CHARACTERISTICS OF THE WORK

The purpose of this research is to develop a framework for restoring images with distorted data using program and mathematical tools, enhancing the accuracy and reliability of image restoration.

In the contemporary digital era, images have become an indispensable medium for information exchange, serving critical roles in fields ranging from medical diagnostics and remote sensing to social media communication and forensic investigation. However, the integrity of digital images is frequently compromised during acquisition, transmission, and storage processes, leading to various forms of degradation including noise corruption, blur, and missing data regions. The restoration of images with missing or corrupted data represents a fundamental challenge in digital image processing, hence the relevance of this topic.

The object of the study is software and mathematical tools for detecting image distortions. The subject of the research includes statistical hypothesis tests, regression tree methods, various distortion types.

For the purposes of this research, it is assumed that the following tasks will be solved:

1. To analyze and evaluate existing methods for image restoration.

This phase includes the comprehensive study of existing image restoration methods, including Fourier filter approaches, statistical hypothesis testing methods and machine learning-based approaches.

2. To develop a unified mathematical framework that integrates different restoration approaches.

At this stage, the task is to implement various algorithms for image restoration with distorted data.

3. To conduct experiments and analysis.

All the above tasks are designed to achieve the main goal of this research, which is to provide a reliable, interpretable, and training-free statistical framework for image distortion detection, and to demonstrate its superiority over regression tree methods.

Provisions for defense:

1. Majority voting rule combining KS test, t-test and 95% confidence interval enables robust distortion detection with Type I error below 2%, outperforming single-test methods.

2. Comparative analysis, results, and conclusions based on experiments with different types of distortions and sample sizes.

3. Software implementation of the methods.

The topic of the dissertation corresponds to the list of priority areas defined by the Decree of the President of the Republic of Belarus dated 01.04.2025 (No. 135) “On priority areas of scientific, scientific-technical and innovative activities for 2026-2030” (area “Digital technologies and artificial intelligence”).

The main provisions and results of the research were discussed at various conferences:

Information Technologies and Systems 2025 (ITS 2025): Proceedings of the international scientific conference, Minsk, November 19, 2025, BSUIR;

Technical Means of Information Security (TSZI): Proceedings of the XXIV International Scientific and Technical Conference, Minsk, April 8, 2026, BSUIR.

SUMMARY OF THE WORK

The focus of this work is to research and compare statistical hypothesis testing methods and regression tree methods for image distortion detection and quality assessment, in order to determine which approach provides more reliable and interpretable results under different distortion types and sample sizes.

The focus of this research is to propose a unified statistical framework based on the Kolmogorov–Smirnov test, *t*-test, and 95% confidence interval combined (a majority voting rule), and to systematically evaluate its performance against a regression tree method that uses only first-and second-order moment features.

The statistical method uses a majority voting rule where at least two of the three tests (KS test, t-test, confidence interval) must conclude "no significant difference" to declare that a distorted image retains statistical equivalence to the original. This approach, requires no training, and provides interpretable outputs including *p*-values, confidence intervals, and normality test results.

The regression tree method is trained exclusively on defect rates ≤0.30 using six features (original image mean μ_o, original image variance σ_o^2, distorted image mean μ_d, distorted image variance σ_d^2, absolute mean difference $|\mu_o - \mu_d|$, variance ratio σ_d^2 / σ_o^2) and employs a fixed decision threshold of 0.8. Although it is computationally faster, it suffers from poor generalization, instability across sample sizes, and lack of interpretability.

Finally, extensive experiments were conducted on two natural images, covering four distortion types (Gaussian noise, salt-and-pepper noise, occlusion, blur), five sample sizes (10, 50, 100, 200, 500 pixels), and defect rates ranging from 0.01 to 0.70. The statistical method achieved a 70% win rate, maintained a maximum tolerable defect rate of 0.70 for global distortions across all sample sizes, and exhibited a Type I error below 2%. In contrast, the regression tree showed severe instability and poor generalization to unseen images. These results demonstrate that the statistical hypothesis testing framework is superior in accuracy, stability, interpretability, and generalization capability, making it the recommended choice for high-reliability image quality assessment applications.

CONCLUSION

This research systematically investigated and compared two fundamentally different approaches for image distortion detection and quality assessment: a statistical hypothesis testing framework and a regression tree method. The statistical framework integrates the Kolmogorov-Smirnov test, t-test, and the 95% confidence interval for the mean difference, combined with a majority voting rule.

The Anderson-Darling (AD) and Shapiro-Wilk (SW) tests are not directly included in the majority voting rule, they play an essential diagnostic role in the overall analysis. For each pair of pixel samples (original and distorted), these tests are applied to assess whether the pixel intensities follow a normal distribution.

The regression tree was designed to use only first- and second-order moment features, trained on defect rates ≤ 0.30, with a fixed decision threshold of 0.8.

Extensive experiments were conducted on four representative distortion types (Gaussian noise, salt-and-pepper noise, occlusion, and blur) across five sample sizes (10, 50, 100, 200, 500 pixels). The results consistently demonstrate that the statistical method significantly outperforms the regression tree in detection accuracy (winning 70% of experimental conditions), stability (Type I error below 2%), interpretability (providing p-values, confidence intervals, and normality test results), and generalization capability (requiring no training and working directly on unseen images). Moreover, the statistical method maintains a maximum tolerable defect rate of 0.70 for Gaussian noise, salt-and-pepper noise, and blur across all sample sizes, while the regression tree exhibits erratic behavior, severe instability, and poor generalization – especially when applied to new image content.

The regression tree offers a four-fold speed advantage (10.1 seconds per image vs. 40.2 seconds for the statistical method), but this comes at the cost of substantially lower accuracy, reliability, and interpretability. Therefore, for high-reliability applications such as medical imaging, remote sensing, and forensic analysis, the statistical method is unequivocally recommended. For real-time applications, a hybrid strategy (regression tree screening followed by statistical verification) may be considered, though it requires further experimental validation.

Future research could extend the statistical framework to multi-channel color images and explore adaptive majority voting rules for specific distortion types. The findings of this research provide clear methodological guidance for selecting appropriate image quality assessment tools in practical scenarios and lay a solid foundation for automated, training-free distortion detection systems.

LIST OF PUBLISHED WORKS

1-A. Jing, W. Statistical analysis of pixel distribution for image distortion detection. / J. Wang, Yu. German // Information Technologies and Systems 2025 (ITS 2025): Proceedings of the International Scientific Conference, Minsk, November 19, 2025 / BSUIR. – Minsk, 2025. – P. 259 – 260.

2-A. Jing, W. Development of mathematical tools for image restoration with distorted data. / J. Wang // Technical Means of Information Security: Proceedings of the XXIV International Scientific and Technical Conference, Minsk, April 8, 2026 / BSUIR [et al.]. - Minsk, 2026. - P. 124 -126.