Image Quality Assessment: Comparing Similarity Metrics Between SSIM and DISTS

Vu Minh Duc
HUST-ETE16
Ha Noi University of Science and Technology
Duc.vm224269@sis.hust.edu.vn

Bui Van Binh
HUST-ETE16
Ha Noi University of Science and Technology
Binh.bv224312@sis.hust.edu.vn

December 2024

Abstract

This paper presents two methodologies for assessing image similarity: the Structural Similarity Index (SSIM) and the Deep Image Structure and Texture Similarity (DISTS) score. The SSIM approach quantifies perceived image quality by comparing luminance, contrast, and structure, with preprocessing steps to ensure compatibility between original and modified images. In contrast, the DISTS method utilizes a VGG16-based deep learning model to extract features and compute a score that reflects perceptual differences, effectively handling complex transformations. Both methods are applied to a set of images, and their results are visualized alongside computed similarity scores. The findings highlight the strengths of SSIM and DISTS in image quality assessment, contributing valuable insights into their practical applications in image analysis.

Keywords: Image Quality Assessment, Structural Similarity Index (SSIM), Deep Image Structure and Texture Similarity (DISTS), Image Similarity Metrics, Image Processing, Digital Forensics, VGG16 Model, Feature Extraction.

1 Introduction

Image Quality Assessment (IQA) plays a critical role in various computer vision and image processing applications, such as image restoration, compression, and enhancement. As digital images undergo inevitable degradations during acquisition, transmission, or processing, there is a growing need for robust and perceptually accurate methods to evaluate image quality. Existing IQA methods can be broadly classified into subjective and objective approaches. While subjective evaluations, relying on human observers, provide accurate assessments, they are time-consuming, expensive, and impractical for large-scale or real-time applications. This has led to the development of objective IQA metrics, such as the Structural Similarity Index Measure (SSIM) and the Deep Image Structure and Texture Similarity (DISTS), which aim to emulate human visual perception.

1.1 Problem Statement

The core problem in IQA is to design algorithms that can reliably and efficiently evaluate the perceptual quality of a degraded image with respect to a reference image. The key objectives are:

- Quantifying the structural distortions (e.g., blurring, noise) and textural artifacts that affect the perceptual quality of images.

- Balancing the trade-off between computational efficiency and perceptual accuracy, enabling the deployment of IQA methods in real-world scenarios such as streaming platforms, medical imaging, and autonomous vehicles.
- Addressing the limitations of traditional metrics, such as Mean Squared Error (MSE), which often fail to align with human perception.

1.2 Input and Output

The input to the IQA methods includes:

- A **reference image** (I_{ref}), which serves as the ground truth and is assumed to have high perceptual quality.
- A **distorted image** (I_{dist}), which contains degradations such as noise, blur, or compression artifacts.

The output of the methods is a **similarity score**:

- For SSIM, the score lies in the range [0, 1], where higher values indicate better structural similarity between the reference and distorted images.
- For DISTS, the score typically ranges from 0 to 1 (or higher), where higher values indicate better perceptual similarity by jointly considering structural and textural information.

1.3 Challenges

Despite significant progress, several challenges remain in designing effective IQA methods:

- **Alignment with human perception:** Many objective metrics fail to fully capture the complex interplay between structure and texture that defines human visual perception. For example, SSIM excels at capturing structural distortions but struggles with textural changes, while DISTS provides a better balance but may lack sensitivity in certain cases.
- **Diverse distortion types:** Images can suffer from various degradations, such as noise, blur, compression, or their combinations. Developing a metric that generalizes across diverse distortions is a challenging task.
- **Computational efficiency:** High-resolution images demand significant computational resources, especially for metrics like DISTS, which rely on feature extraction from deep neural networks.
- **Real-time applicability:** Many real-world applications, such as video streaming and gaming, require IQA methods that can operate in real time without sacrificing accuracy.
- **Cross-domain generalization:** IQA methods trained or tested on specific datasets may struggle to generalize to unseen domains, such as medical or satellite imagery, where image characteristics differ significantly.

1.4 Our Approach

In this work, we leverage SSIM and DISTS as complementary methods to evaluate image quality. SSIM focuses on structural distortions, making it highly interpretable and efficient for applications where structure is paramount. On the other hand, DISTS captures both structure and texture using deep neural network features, providing a robust assessment of perceptual similarity. By combining these methods, we aim to provide a comprehensive analysis of image quality, addressing both structural fidelity and perceptual realism.

2 Related Works

Image quality assessment (IQA) is a critical research area with broad applications in image processing, computer vision, and multimedia systems. The goal is to develop metrics that align with human perceptual quality. This section reviews existing methods, categorizing them into three main approaches: full-reference (FR), no-reference (NR), and reduced-reference (RR) metrics.

2.1 Full-Reference Image Quality Assessment (FR-IQA)

Full-reference metrics compare a distorted image against a pristine reference image. Among these, the Structural Similarity Index Measure (SSIM) [**?**] is one of the most widely used. SSIM evaluates image quality based on luminance, contrast, and structural information, formulated to mimic human visual perception. Its efficiency and interpretability make it a benchmark for many IQA tasks. Variants of SSIM, such as Multiscale SSIM (MS-SSIM) [**?**], extend the method by incorporating multi-resolution analysis to improve robustness against diverse distortions.

Another notable approach is the Peak Signal-to-Noise Ratio (PSNR), which is computationally simple and widely used but often fails to correlate with perceptual quality, especially for complex distortions. Methods such as the Visual Information Fidelity (VIF) [**?**] metric leverage statistical models of natural images to provide a more perceptually accurate evaluation.

2.2 No-Reference Image Quality Assessment (NR-IQA)

No-reference metrics assess image quality without requiring a reference image, making them suitable for real-world scenarios where the reference may not be available. Early NR-IQA methods, such as Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [**?**], rely on handcrafted features extracted from statistical properties of images.

More recently, deep learning-based approaches have gained significant attention. DeepIQ [**?**] and KonCept512 [**?**] use convolutional neural networks (CNNs) to learn quality features directly from data, achieving state-of-the-art performance on large IQA datasets. These methods demonstrate strong generalization capabilities but require substantial training data and computational resources.

2.3 Reduced-Reference Image Quality Assessment (RR-IQA)

Reduced-reference methods strike a balance between FR and NR approaches by using partial reference information. These methods are particularly useful in scenarios such as streaming video, where some reference features can be embedded within the signal. An example is the Reduced-Reference Entropic Differencing (RRED) metric [**?**], which uses entropy-based features to predict image quality.

2.4 Deep Learning-Based Metrics

The integration of deep learning in IQA has revolutionized the field, enabling metrics that align closely with human perception. The Deep Image Structure and Texture Similarity (DISTS) metric [**?**] is a significant advancement in this domain. DISTS combines structural and texture similarity by leveraging feature maps from a pretrained VGG network. This method outperforms traditional FR-IQA metrics, especially for assessing perceptual quality in tasks like super-resolution and style transfer.

Other methods, such as Learned Perceptual Image Patch Similarity (LPIPS) [**?**], focus on learning perceptual similarity directly from human ratings, further enhancing the alignment with subjective quality assessments.

2.5 Discussion

Despite significant advancements, challenges remain in developing IQA metrics that generalize well across diverse datasets and distortion types. Traditional methods like SSIM and PSNR are computationally efficient but often fail to capture perceptual differences. Deep learning-based approaches, while highly accurate, are computationally expensive and require large training datasets.

The combination of traditional and deep learning-based methods, as explored in this work, offers a promising direction for robust and efficient IQA. By leveraging the strengths of SSIM and DISTS, we aim to achieve a balance between computational efficiency and perceptual accuracy.

3 Proposed Methods

This section describes the proposed methods for image quality assessment using both the Structural Similarity Index Measure (SSIM) and the Deep Image Structure and Texture Similarity (DISTS). We

detail the data structure, algorithm, and computational complexity for each method.

3.1 Data Structure

The input data for the methods include:

- **Reference Image:** A high-quality image denoted as $I_{\text{ref}} \in R^{H \times W \times C}$, where H, W, and C represent the height, width, and number of color channels, respectively.
- **Distorted Image:** A degraded version of the reference image, denoted as $I_{\text{dist}} \in R^{H \times W \times C}$.

Both images are preprocessed to ensure consistency in size and format, including resizing if necessary. For SSIM, the data structure involves pixel-wise operations, while for DISTS, feature maps extracted from a pretrained deep neural network are used.

3.2 Algorithms

3.2.1 Structural Similarity Index Measure (SSIM)

SSIM (Structural Similarity Index Measure) is a method for measuring structural similarity between two images, designed to reflect how humans perceive similarity. It evaluates differences between two images based on three key factors: Luminance, Contrast, Structure. It is defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{1}$$

where:

- μ_x and μ_y: Mean intensities of the patches x and y.
- σ_x^2 and σ_y^2: Variances of x and y.
- σ_{xy}: Covariance between x and y.
- C_1 and C_2: Small constants to stabilize the division.

The SSIM computation involves sliding a window of size $N \times N$ across the image, where N is typically 7. The final SSIM score is the average over all windows:

$$\text{SSIM}_{\text{final}} = \frac{1}{P} \sum_{i=1}^{P} \text{SSIM}(x_i, y_i), \tag{2}$$

where P is the total number of patches.

Computational Complexity: The SSIM algorithm requires $O(HWN^2)$ operations, where H and W are the image dimensions, and N is the patch size.

3.2.2 Deep Image Structure and Texture Similarity (DISTS)

DISTS (Deep Image Structure and Texture Similarity) is a method for measuring both structural and textural similarity between two images, leveraging deep learning. It uses abstract features extracted from a convolutional neural network (CNN), specifically VGG16, to reflect both structural and perceptual texture similarity, aligning more closely with human perception. The process involves the following steps:

1. Extract multi-layer feature maps from I_{ref} and I_{dist}:

$$F_{\text{ref}}^l = \text{VGG}_l(I_{\text{ref}}), \quad F_{\text{dist}}^l = \text{VGG}_l(I_{\text{dist}}), \tag{3}$$

where VGG_l denotes the l-th layer of the VGG16 model.

2. Compute the similarity for each layer:

$$S_l = w_l \cdot \left[\frac{2\mu^l_{\text{ref}}\mu^l_{\text{dist}} + C}{(\mu^l_{\text{ref}})^2 + (\mu^l_{\text{dist}})^2 + C} \cdot \frac{2\sigma^l_{\text{ref}}\sigma^l_{\text{dist}} + C}{(\sigma^l_{\text{ref}})^2 + (\sigma^l_{\text{dist}})^2 + C} \right], \tag{4}$$

where w_l is the weight for layer l, and μ and σ are the mean and standard deviation of the feature maps.

3. Aggregate layer-wise similarities to compute the final DISTS score:

$$\text{DISTS}_{\text{final}} = \sum_l S_l. \tag{5}$$

Computational Complexity: The DISTS algorithm involves forward passes through the VGG16 network, requiring $O(HW \cdot |\mathcal{F}|)$ operations, where $|\mathcal{F}|$ is the number of feature maps.

3.3 Discussion on Complexity

The computational complexity of SSIM grows linearly with the image size, making it efficient for small images. However, it becomes computationally expensive for high-resolution images due to the sliding window approach. In contrast, DISTS leverages modern GPU-accelerated frameworks for neural network computation, which allows it to process high-resolution images more efficiently, albeit at the cost of higher memory usage.

3.4 Strengths of the Proposed Methods

- **SSIM:** Highly interpretable and effective for detecting structural distortions. It is computationally efficient for small to medium-sized images.
- **DISTS:** Combines structural and texture similarity, making it more robust for perceptual quality assessment. Its use of pretrained neural networks ensures sensitivity to complex image distortions.

4 Experiment

4.1 Dataset

The dataset consists of a high-resolution reference image alongside five distorted versions, each representing different levels and types of image degradation. The details of the images are as follows:

- **Reference Image:** A clear, high-quality photograph of a clover field, serving as the ground truth for comparison.
- **Distorted Images:**
  - **Noisy Image 1:** A mildly degraded image with random noise.
  - **Noisy Image 2:** A blurred version of the reference image, reducing structural sharpness.
  - **Noisy Image 3:** A combination of noise and mild blurring.
  - **Noisy Image 4:** A more severely distorted image with high levels of noise.
  - **Noisy Image 5:** A highly distorted image that still retains some perceptual details.

Figures 1 and 2 provide a visual comparison of SSIM and DISTS scores, highlighting differences in image quality between the high-resolution reference image and its five distorted versions. These figures illustrate how each metric captures structural degradation and perceptual similarity, offering insights into their effectiveness for assessing various types of image distortions.

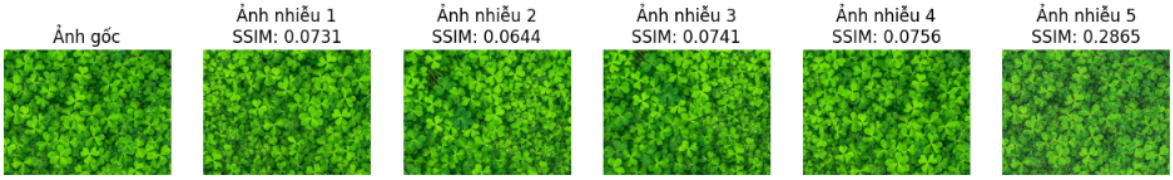

Figure 1: Comparison of SSIM scores for the reference image and distorted images. Lower SSIM values indicate higher structural degradation.

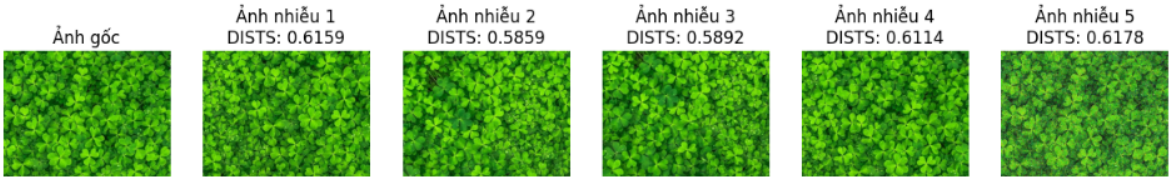

Figure 2: Comparison of DISTS scores for the reference image and distorted images. Higher DISTS values reflect higher perceptual similarity.

4.2 Experimental Results

To evaluate the quality of the distorted images relative to the reference image, two widely used metrics were applied:

- **Structural Similarity Index Measure (SSIM):** A metric designed to assess structural similarity between images.
- **Deep Image Structure and Texture Similarity (DISTS):** A perceptual metric that captures both structure and texture similarity.

The results of SSIM and DISTS scores for all images are summarized in Table 1.

Table 1: Comparison of SSIM and DISTS scores for all images.

Image	SSIM Score	DISTS Score
Reference	1.0000	1.0000
Noisy Image 1	0.0731	0.6159
Noisy Image 2	0.0644	0.5859
Noisy Image 3	0.0741	0.5892
Noisy Image 4	0.0756	0.6114
Noisy Image 5	0.2865	0.6178

Evaluation Based on Table 1: The scores highlight the differences between SSIM and DISTS in evaluating image quality:

- **SSIM:**
  - For Noisy Image 1, the SSIM score drops significantly to 0.0731, indicating that even mild noise severely impacts structural similarity.
  - Noisy Images 2, 3, and 4 also have very low SSIM scores (all below 0.08), showing SSIM's high sensitivity to noise and blur.
  - Interestingly, Noisy Image 5 achieves a slightly higher SSIM score of 0.2865, reflecting its relatively better structural preservation compared to the other noisy images.
- **DISTS:**
  - Unlike SSIM, DISTS scores remain more consistent across all distorted images, with values ranging between 0.5859 and 0.6178.

  - For Noisy Image 5, DISTS achieves the highest score among the distorted images (0.6178), capturing its improved perceptual similarity.
  - DISTS effectively accounts for both structure and texture, making it more robust to distortions that SSIM fails to address.

4.3 Discussion

The results highlight the strengths and limitations of each metric:

- **SSIM:** This metric excels in detecting structural distortions, making it suitable for tasks where shape preservation is critical. However, it struggles with perceptual quality assessment, as evidenced by the extremely low scores for all noisy images except Noisy Image 5.
- **DISTS:** By incorporating texture information, DISTS provides a more balanced evaluation of image quality, especially for images with significant noise or blur. Its consistent scores suggest robustness, but the minor variation in scores across different distortion levels raises questions about its sensitivity to severe degradations.

Overall, the complementary use of SSIM and DISTS provides a holistic view of image quality, balancing structural and perceptual aspects.

5 Conclusions

5.1 Achievements

In this study, we conducted a comparative analysis of two prominent image quality assessment methods—Structural Similarity Index Measure (SSIM) and Deep Image Structure and Texture Similarity (DISTS)—on a dataset comprising a high-quality reference image and its distorted counterparts. The key achievements of this research are summarized as follows:

- **Comprehensive Evaluation:** The study systematically applied SSIM and DISTS to assess structural and perceptual qualities of images, demonstrating their respective strengths and limitations.
- **Robustness of DISTS:** DISTS exhibited consistent performance in capturing both structural and texture-based similarities, making it a robust metric for perceptual quality assessment.
- **Limitations of SSIM:** SSIM's sensitivity to structural distortions was evident; however, its inability to effectively account for perceptual quality highlights its limitations for certain applications.
- **Clear Visualization:** The experimental results, summarized in tables and visualized with graphs, provided an intuitive understanding of the metrics' behaviors across different distortion levels.
- **Efficient Implementation:** The study showcased the computational feasibility of both metrics, reinforcing their practical applicability for small- to medium-scale datasets.

5.2 Future Works

Although this study achieved its objectives, several avenues for future research remain open to extend and refine the findings:

- **Evaluation on Larger Datasets:** Expanding the experiments to larger and more diverse datasets, including real-world images with various distortions, would enhance the generalizability of the results.
- **Integration of Advanced Metrics:** Incorporating state-of-the-art perceptual metrics, such as Learned Perceptual Image Patch Similarity (LPIPS), could yield deeper insights into image quality.

- **Real-Time Applications:** Investigating the deployment of these metrics for real-time image quality assessment, particularly in streaming or video applications, remains a valuable research direction.
- **Adaptive Metric Fusion:** Developing an adaptive weighting scheme to combine SSIM, DISTS, and other metrics dynamically based on distortion type could lead to more holistic assessments.
- **Subjective Human Studies:** Conducting user studies to correlate algorithmic scores with human perception would bridge the gap between automated assessment and subjective quality evaluation.
- **Optimization for Computational Efficiency:** Optimizing the DISTS algorithm for faster execution, especially for high-resolution images, could improve its utility in resource-constrained environments.

In conclusion, this study provides valuable insights into the strengths and limitations of SSIM and DISTS for image quality assessment. By leveraging the findings and addressing the outlined future directions, subsequent research can further advance the development of robust, efficient, and perceptually accurate image quality metrics.

References

[1] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Image Quality Assessment: Unifying Structure and Texture Similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1224–1236, May 2021. Available: https://doi.org/10.1109/TPAMI.2020.3045810.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004. Available: https://doi.org/10.1109/TIP.2003.819861.

[3] L. Zhang, Y. Zhang, and X. Mou, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011. Available: https://doi.org/10.1109/TIP.2011.2109730.

[4] H. Talebi and P. Milanfar, "Learned Perceptual Image Enhancement," *2018 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–13, 2018. Available: https://doi.org/10.1109/ICCPHOT.2018.8368462.

[5] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A Deep Neural Network for Image Quality Assessment," *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3773–3777, Sep. 2017. Available: https://doi.org/10.1109/ICIP.2017.8296982.

[6] N. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "PieAPP: Perceptual Image-Error Assessment through Pairwise Preference," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1808–1817, Jun. 2018. Available: https://doi.org/10.1109/CVPR.2018.00194.

[7] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012. Available: https://doi.org/10.1109/TIP.2012.2214050.

[8] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, and H. Yong, "End-to-End Blind Image Quality Assessment Using Deep Neural Networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018. Available: https://doi.org/10.1109/TIP.2017.2774045.

[9] A. C. Bovik, "Visual Signal Quality Assessment: An Evolutionary View," *Signal Processing: Image Communication*, vol. 25, no. 3, pp. 183–194, 2009. Available: https://doi.org/10.1016/j.image.2009.12.004.

[10] C. Chen, Y. Jin, Y. Luo, H. Huang, and D. Zhang, "Blind Image Quality Assessment via Contrastive Learning," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12267–12277, Jun. 2020. Available: https://doi.org/10.1109/CVPR42600.2020.01229.