Full-Reference Image Quality Metrics:
Classification and Evaluation
Full-Reference Image Quality Metrics: Classification and Evaluation

Marius Pedersen
Gjøvik University College
N-2802 Gjøvik
Norway
marius.pedersen@hig.no

Jon Yngve Hardeberg
Gjøvik University College
N-2802 Gjøvik
Norway
jon.hardeberg@hig.no

now
the essence of knowledge

Boston – Delft
Editorial Scope

Foundations and Trends® in Computer Graphics and Vision will publish survey and tutorial articles in the following topics:

- Rendering: Lighting models; Forward rendering; Inverse rendering; Image-based rendering; Non-photorealistic rendering; Graphics hardware; Visibility computation
- Shape: Surface reconstruction; Range imaging; Geometric modelling; Parameterization;
- Mesh simplification
- Animation: Motion capture and processing; Physics-based modelling; Character animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape Representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and Video Retrieval
- Video analysis and event recognition
- Medical Image Analysis
- Robot Localization and Navigation

Information for Librarians
Foundations and Trends® in Computer Graphics and Vision, 2011, Volume 7, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.
Full-Reference Image Quality Metrics: Classification and Evaluation

Marius Pedersen\textsuperscript{1} and Jon Yngve Hardeberg\textsuperscript{2}

\textsuperscript{1} Gjøvik University College, Norwegian Color Research Laboratory, P.O. Box 191, N-2802 Gjøvik, Norway, marius.pedersen@hig.no
\textsuperscript{2} Gjøvik University College, Norwegian Color Research Laboratory, P.O. Box 191, N-2802 Gjøvik, Norway, jon.hardeberg@hig.no

Abstract

The wide variety of distortions that images are subject to during acquisition, processing, storage, and reproduction can degrade their perceived quality. Since subjective evaluation is time-consuming, expensive, and resource-intensive, objective methods of evaluation have been proposed. One type of these methods, image quality (IQ) metrics, have become very popular and new metrics are proposed continuously. This paper aims to give a survey of one class of metrics, full-reference IQ metrics. First, these IQ metrics were classified into different groups. Second, further IQ metrics from each group were selected and evaluated against six state-of-the-art IQ databases.
# Contents

1 Introduction 1

2 Classification of Image Quality Metrics 3

2.1 Existing Classification of Image Quality Metrics 3

2.2 Proposal for Classification of Image Quality Metrics 5

3 Survey of Image Quality Metrics 7

3.1 Mathematically Based Metrics 8

3.2 Low-level Metrics 13

3.3 High-level Metrics 19

3.4 Other Approaches 24

4 Evaluation of Image Quality Metrics 29

4.1 TID2008 Database, Ponomarenko et al. 32

4.2 Luminance Changed Images, Pedersen et al. 38

4.3 JPEG and JPEG2000 Compressed Images, Caracciolo 40

4.4 IVC Database, Le Callet et al. 43

4.5 Images Altered in Contrast, Lightness, and Saturation, Ajagamelle et al. 45

4.6 Gamut Mapped Images, Dugay et al. 47

4.7 Overall Performance of the Metrics 50
Advances are rapid in the imaging industry, and new and more advanced products are continuously being introduced into the market. In order to verify that the new technologies produce higher-quality images than the current technology, some kind of quality assessment is required.

There are two main methods of assessing image quality (IQ): subjective or objective. The first is carried out by human observers, while the second does not involve observers. To make an objective assessment, one can use measuring devices to obtain numerical values; another method is to use IQ metrics. These IQ metrics are usually developed to take into account the human visual system (HVS), and thus have the goal of correlating with subjective assessment.

Many IQ metrics have been proposed in the literature; a brief summary of more than 100 metrics was given by Pedersen and Hardeberg [125]. These metrics stem from different ideas, and they have been made for different purposes, such as to quantify distortions, produce benchmarks, monitor quality, optimize a process, or indicate problem areas. Because different metrics have different goals, it is important to keep in mind their areas of use when evaluating their performance.
Existing surveys, such as the one by Wang and Bovik [166], mostly focus on grayscale IQ metrics whereas the survey by Avcibas et al. [6] covers only simple statistical metrics. In this survey, we continue the work started in Pedersen and Hardeberg [125] and we carry out a comprehensive survey and evaluation of color and grayscale IQ metrics.

Our goal is to classify IQ metrics into separate groups and to evaluate their correspondence with the percept. Such a classification can be used to select the most appropriate IQ metric for a given problem or distortion. It will also provide a better understanding of the state-of-the-art of IQ metrics, which can be used to improve or develop new metrics that correlate better with the percept. Because the original is available in many situations, we limit our survey to full-reference IQ metrics, where both the complete original and reproduction are used for the calculation of quality. Additionally, more work has been carried out on full-reference IQ metrics than on reduced-reference and no-reference metrics. The two latter types of IQ metrics are also considered to be more difficult to assess than full-reference metrics [170].

In the literature many different terms have been used, such as IQ, image difference, image fidelity, and image similarity. If possible we will use the general term IQ. In addition, we use the term metric as a general term even though not all metrics fulfill the requirements to be a metric in mathematical terms. Other terms, such as index, measure, and criterion, have also been used in the literature.

The survey is organized as follows: first, we classify IQ metrics into groups; then we go through IQ metrics within each group. This is followed by evaluations of selected IQ metrics from each group. Finally, we report our conclusions.
References


References


Full text available at: http://dx.doi.org/10.1561/0600000037


[34] CIE, “Industrial colour-difference evaluation,” publication CIE 116-95, bureau central de la CIE, 1995.
References


References


References


Full text available at: http://dx.doi.org/10.1561/0600000037
References

References


References


References


References


References


