INTERNATIONAL STANDARD

ISO 20462-3

Second edition
2012-05-15

Photography — Psychophysical experimental methods for estimating image quality —

Part 3: Quality ruler method

Photographie — Méthodes psychophysiques expérimentales pour estimer la qualité d'image —

Partie 3: Méthode de la règle de qualité

Reference number
ISO 20462-3:2012(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 20462-3 was prepared by Technical Committee ISO/TC 42, *Photography*.

This second edition cancels and replaces the first edition (ISO 20462-3:2005), which has been technically revised.

ISO 20462 consists of the following parts, under the general title *Photography — Psychophysical experimental methods for estimating image quality*:

— *Part 1: Overview of psychophysical elements*

— *Part 2: Triplet comparison method*

— *Part 3: Quality ruler method*

Introduction

There are many circumstances under which it is desirable to quantify image quality in a standardized fashion that facilitates interpretation of results within a given experiment and/or comparison of results between different experiments. Such information can be of value in assessing the performance of different capture or display devices, image processing algorithms, etc. under various conditions. However, the choice of the best psychometric method for a particular application may be difficult to make, and interpretation of the rating scales produced by the numerical analyses is frequently ambiguous. Furthermore, none of the commonly used rating techniques provides an efficient mechanism for calibration of the results against a standardized numerical scale or associated physical references, which is desirable when results of different experiments are to be compared or integrated.

ISO 20462-1, ISO 20462-2 and this part of ISO 20462 address the need for documented means of determining image quality in a calibrated fashion. ISO 20462-1 provides an overview of practical psychophysics and aids in identifying the better choice between the two alternative approaches described in ISO 20462-2 (triplet comparison method[2][3][4]) and this part of ISO 20462 (quality ruler method[5]). These two techniques are complementary and together are sufficient to span a wide range of practical applications. ISO 20462-2 and this part of ISO 20462 document both specific experimental methods and associated data reduction techniques. It is the intent of these methods to produce results that are not merely directional in nature, but are expressed in terms of relative or fixed scales that are calibrated in terms of just noticeable differences (JNDs), so that the significance of experimentally measured stimulus differences is readily ascertained.

The quality ruler method described in this part of ISO 20462 is particularly suitable for measuring quality differences exceeding one JND. The ratings given by an observer can be converted to JND values in real time, rather than having to wait until the entire experimental data set has been collected and analysed. Furthermore, with suitable reference stimuli, the quality ruler method permits the results to be reported using the standard quality scale (SQS), a fixed numerical scale that:

a) is anchored against physical standards;

b) has one unit corresponding to one JND; and

c) has a zero point corresponding to an image having little identifiable information content.

Reflection prints calibrated against the absolute SQS, which are referred to as standard reference stimuli (SRS), will be available at the Standards Resources link at www.imaging.org. Digital Reference Stimuli (DRS) will also be provided at the Standards Resources link at www.imaging.org. These images, when displayed on a high-quality monitor and viewed correctly, will have approximately known absolute SQS values, and accurately known relative SQS values (JNDs). Included with the images will be software for running softcopy quality ruler experiments. This part of ISO 20462 also describes how users can conveniently generate their own quality ruler images with correct relative calibrations and, if desired, calibrate them absolutely against the SRS.

The International Organization for Standardization (ISO) draws attention to the fact that it is claimed that compliance with this document may involve the use of US Patent Numbers 6,639,999 and 6,658,139 concerning the quality ruler given in Clauses 4 to 6.

ISO takes no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has ensured ISO that he is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with ISO. Patent inquiries may be addressed to:

General Council and Senior Vice President

Eastman Kodak Company

345 State Street

Rochester, NY 14650

USA

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those identified above. ISO shall not be held responsible for identifying any or all such patent rights.

Photography — Psychophysical experimental methods for estimating image quality —

Part 3:
Quality ruler method

1 Scope

This part of ISO 20462 specifies:

a) the nature of a quality ruler;

b) hardcopy and softcopy implementations of quality rulers;

c) how quality rulers may be generated or obtained; and

d) the standard quality scale (SQS), a fixed numerical scale that may be measured using quality rulers.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3664, *Graphic technology and photography — Viewing conditions*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
artefactual attribute
attribute of image quality that, when evident in an image, nearly always leads to a loss of overall image quality

EXAMPLES Noise, aliasing.

NOTE The commonly used terms “defect” and “impairment” are similar in meaning.

3.2
attribute
aspect, dimension, or component of overall image quality

cf. **artefactual attribute** (3.1) and **preferential attribute** (3.11)

EXAMPLES Image structure properties such as sharpness and noise; colour and tone reproduction properties such as contrast, colour balance, and relative colourfulness; digital artefacts such as aliasing, contouring, and compression defects.

3.3
digital reference stimuli
DRS
set of digital images used in the softcopy ruler, which vary in sharpness and are calibrated against the standard quality scale (SQS) when suitably displayed and viewed

NOTE The DRS will be available at the Standards Resources link at www.imaging.org.

3.4
image quality
impression of the overall merit or excellence of an image, as perceived by an observer neither associated with the act of photography nor closely involved with the subject matter depicted

NOTE The purpose of defining image quality in terms of third-party (uninvolved) observers is to eliminate sources of variability that arise from more idiosyncratic aspects of image perception and pertain to attributes outside the control of imaging system designers.

3.5
instructions
set of directions given to the observer for performing the psychophysical evaluation task

3.6
just noticeable difference
JND
stimulus difference that leads to a 75:25 proportion of responses in a paired comparison task

cf. **quality JND** (3.13)

3.7
magnitude estimation method
psychophysical method involving the assignment of a numerical value to each test stimulus that is proportional to image quality; typically, a reference stimulus with an assigned numerical value is present to anchor the rating scale

NOTE The numerical scale resulting from a magnitude estimation experiment is usually assumed to constitute a ratio scale which, ideally, is a scale in which a constant percentage change in value corresponds with one JND. In practice, modest deviations from this behaviour occur, complicating the transformation of the rating scale into units of JNDs without inclusion of unidentified reference stimuli (having known quality) among the test stimuli.

3.8
multivariate
⟨series of test or reference stimuli⟩ varying in multiple attributes of image quality

3.9
observer
individual performing the subjective evaluation task in a psychophysical method

3.10
paired comparison method
psychophysical method involving the choice of which of two simultaneously presented stimuli exhibits greater or lesser image quality or an attribute thereof, in accordance with a set of instructions given to the observer

NOTE 1 Two limitations of the paired comparison method are as follows.

a) If all possible stimulus comparisons are done, as is usually the case, a large number of assessments are required for even modest numbers of experimental stimulus levels [if n levels are to be studied, $n\,(n-1)/2$ paired comparisons are needed].

b) If a stimulus difference exceeds approximately 1,5 JNDs, the magnitude of the stimulus difference cannot be directly estimated reliably because the response saturates as the proportions approach unanimity.

NOTE 2 However, if a series of stimuli having no large gaps are assessed, the differences between more widely separated stimuli may be deduced indirectly by summing smaller, reliably determined (unsaturated) stimulus differences. The standard methods for transformation of paired comparison data to an interval scale (a scale linearly related to JNDs) perform statistically optimized procedures for inferring the stimulus differences, but they may yield unreliable results when saturated responses are included in the analysis.

3.11
preferential attribute
attribute of image quality that is invariably evident in an image, and for which the preferred degree is a matter of opinion, depending upon both the observer and the image content

EXAMPLES Colour and tone reproduction properties such as contrast and relative colourfulness.

NOTE 1 Because the perceived quality associated with a preferential attribute is dependent upon both the observer and image content, in studies involving variations of preferential attributes, particular care is needed in the selection of representative sets of stimuli and groups of observers.

NOTE 2 The term “noticeable” in “just noticeable difference” is not linguistically strictly correct when applied to a preferential attribute, but is nonetheless retained in this part of ISO 20462 for convenience. For example, the higher contrast stimulus of a pair differing only in contrast might be readily identified by all observers, whereas there might be a lack of consensus regarding which of the two images was higher in overall image quality. Nonetheless, if the responses from the paired comparison for quality were in the proportion of 75:25, the image chosen more frequently would be said to be one JND higher in quality. The JND is best regarded as a measurement unit tied to the predicted or measured outcome of a paired comparison.

3.12
psychophysical method
experimental technique for subjective evaluation of image quality or attributes thereof, from which stimulus differences in units of JNDs may be estimated

cf. **magnitude estimation method** (3.7), **paired comparison method** (3.10), **quality ruler method** (3.14), and **triplet comparison** (3.23)

3.13
quality just noticeable difference
quality JND
measure of the significance or importance of quality variations, corresponding to a stimulus difference that leads to a 75:25 proportion of responses in a paired comparison task in which multivariate stimuli pairs are assessed in terms of overall image quality

NOTE See **attribute JND** (3.3) and **quality JND** (3.14) in ISO 20462-1:2005 for greater detail.

3.14
quality ruler method
psychophysical method that involves quality or attribute assessment of a test stimulus against a series of ordered, univariate reference stimuli that differ by known numbers of JNDs

3.15
reference stimulus
image provided to the observer for the purpose of anchoring or calibrating the perceptual assessments of test stimuli in such a manner that the given ratings may be converted to JND units

NOTE The plural is reference stimuli.

3.16
scene
content or subject matter of an image, or a starting image from which multiple stimuli may be produced through different experimental treatments

NOTE Typically, stimuli depicting the same scene are compared in a psychophysical experiment because it is the effect of the treatment that is of interest, and differences in image content could cause spurious effects. In cases where scene content is not matched, a number of scenes should be used so that scene effects may be expected to average out.

3.17
standard quality scale
SQS
fixed numerical scale of quality having the following properties:

a) the numerical scale is anchored against physical standards;

b) a one unit increase in scale value corresponds to an improvement of one JND of quality; and

c) a value of zero corresponds to an image having so little information content that the nature of the subject of the image is difficult to identify.

NOTE SQS_1 (primary SQS) denotes values obtained through assessments traceable to the standard reference stimuli (SRS). SQS_2 (secondary SQS) denotes values obtained through assessments traceable to the digital reference stimuli (DRS) or the average scene relationship (see 7.2).

3.18
standard reference stimuli
SRS
set of reflection prints used in the hardcopy quality ruler, which vary in sharpness and are calibrated against the standard quality scale (SQS)

NOTE The SRS will be available at the Standards Resources link at www.imaging.org.

3.19
stimulus
image presented or provided to the observer either for the purpose of anchoring a perceptual assessment (a reference stimulus) or for the purpose of subjective evaluation (a test stimulus)

NOTE The plural is stimuli.

3.20
suppression
perceptual effect in which one attribute is present in a degree that seriously degrades image quality and thereby reduces the impact that other attributes have on overall quality, compared to the impact they would have had in the absence of the dominant attribute

NOTE To generate reference stimuli that are separated by a specified number of JNDs based on variations in one attribute, it will be necessary to ensure that other attributes do not significantly suppress the impact of the varied attribute.

3.21
test stimulus
image presented to the observer for subjective evaluation

NOTE The plural is test stimuli.

3.22
treatment
controlled or characterized source of the variations between test stimuli (excluding scene content) that are to be investigated in a psychophysical experiment

EXAMPLES Different image processing algorithms, variations in capture or display device properties, changes in image capture conditions (e.g. camera exposure), etc.

NOTE Different treatments may be achieved through hardware or software changes, or may be numerical simulations of such effects. Typically, a series of treatments is applied to multiple scenes, each generating a series of test stimuli. The effect of the treatment may then be determined by averaging the results over scene and observer to improve signal-to-noise and reduce the likelihood of systematic bias.

3.23
triplet comparison
psychophysical method that involves the simultaneous scaling of three test stimuli with respect to image quality or an attribute thereof, in accordance with a set of instructions given to the observer

NOTE The triplet comparison method is described in more detail in ISO 20462-2.

3.24
univariate
⟨series of test or reference stimuli⟩ varying only in a single attribute of image quality

4 Quality ruler experiments

4.1 General properties of quality rulers

A quality ruler is a univariate series of reference stimuli depicting the same scene and having known stimulus differences expressed in JNDs of quality. The reference stimuli are presented to the observer in a fashion facilitating:

a) the identification of the reference stimuli closest in quality to the test stimulus; and

b) the comparison of the test stimulus to those reference stimuli under rigorously matched viewing conditions.

Both hardcopy (Clause 5) and softcopy (Clause 6) implementations of quality rulers are described in this part of ISO 20462. Ruler images may be generated by the user (Clause 7). Reflection prints varying in sharpness and calibrated against the SQS are referred to as standard reference stimuli (SRS) (Clause 8). Analogous digital images, suitable for softcopy display, are referred to as digital reference stimuli (DRS).

The SRS may be used as ruler images or used to calibrate user-generated ruler images on an absolute basis, as distinguished from the relative calibration described in Clause 7.

4.2 Experimental conditions and reported results

Requirements regarding observer selection, test stimulus properties, instructions to the observer, viewing conditions, and reporting of results are set forth in ISO 20462-1.

NOTE 1 Sample instructions to the observer for quality ruler experiments are provided in informative Annex A (hardcopy), informative Annex B (softcopy binary sort paired comparison), and informative Annex F (softcopy slider bar matching). An example of results from quality ruler experiments is provided in informative Annex E.

The viewing requirements of ISO 3664 shall be met, except as modified in ISO 20462-1:2005, 4.4.

Reported values of quality in JNDs or SQS units shall be specifically identified if they are calculated from data 20 % or more of which fall at one of the ends of, or outside, the range of the quality ruler from which they were derived.

NOTE 2 Values based on ratings outside the range of the ruler will be less reliable because of extrapolation effects. In addition, when test samples fall within a JND or two of the high quality end of the ruler, a slight bias may result from observers avoiding use of ratings outside the ruler range. When preferential attributes (e.g. of colour and tone reproduction) are assessed using a quality ruler, it may be desirable to degrade all the test stimuli slightly by blurring (in the case of a ruler varying in sharpness) to allow headroom for test stimuli that are preferred over the reference stimulus.

The pedigree of the rulers used shall be reported, which entails specifying whether they are SRS, DRS, or were otherwise generated. If the latter, the attribute varied in the rulers shall be stated. If such rulers vary in sharpness, the method of calibration shall be stated, which shall either be by comparison with SRS or DRS, or using the average scene relationship (see 7.2).

SQS values determined using the hardcopy SRS, or quality ruler images that have been judged directly against the SRS, and so are rigorously calibrated, shall be denoted as primary SQS (SQS_1) values. SQS values determined using the DRS, or quality ruler images that have been judged against the DRS, or the average scene relationship (see 7.2), and so are less rigorously calibrated, shall be denoted as secondary SQS (SQS_2) values.

4.3 Attributes varied in quality rulers

Clause 7 describes the generation of reference stimuli for rulers varying in sharpness, through modification of the modulation transfer function (MTF) of the system generating the images. Quality rulers may alternatively vary in other attributes, although only one attribute shall change within a given ruler. Alternative attributes that are varied in a quality ruler should be artefactual in nature.

NOTE The variation of preferential attributes within quality rulers is discouraged because of the additional variability associated with such attributes. Sharpness has been selected as the reference attribute because of several desirable characteristics:

a) it is easily manipulated through image processing;

b) it is correlated with MTF, which is readily determinable;

c) it has low scene and observer variability; and

d) it exerts a strong influence on quality in practical imaging systems.

Quality rulers varying in attributes other than sharpness shall be calibrated by having their reference stimuli rated against quality rulers varying in sharpness and meeting the criteria stated in this part of ISO 20462. The calibration experiment shall meet the specifications set forth in ISO 20462-1 and in this part of ISO 20462, with the exception that data from a minimum of 20 observers shall be averaged to determine the calibration.

5 Hardcopy quality ruler implementation

5.1 Physical apparatus

The hardcopy quality ruler apparatus shall consist of the following:

a) a sliding or translating fixture onto or into which a series of reference stimuli may be mounted or inserted (the ruler);

b) a test stimulus fixture in close proximity to the ruler;

c) a base surface upon which the ruler and the test stimulus fixture are attached;

d) an illumination system; and

e) a headrest or other device constraining the viewing distance (the distance from the observer's eye to the test and reference stimuli).

The ruler shall be constructed so that the observer may easily slide it to bring any of two reference stimuli into direct comparison with the test stimulus. In this triangular configuration of one test stimulus and two reference stimuli, the illumination level, illumination angle, viewing distance, and viewing angle shall be sensibly matched between the three stimuli. These features are illustrated in Figure 1.

