TECHNICAL REPORT

ISO/IEC TR 29170-1

First edition
2017-10

Information technology — Advanced image coding and evaluation —

Part 1: Guidelines for image coding system evaluation

Technologies de l'information — Codage d'image avancé et évaluation —

Partie 1: Lignes directices pour l'évaluation des systèmes de codage d'image

Reference number
ISO/IEC TR 29170-1:2017(E)

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

A list of all parts in the ISO/IEC 29170 series can be found on the ISO website.

Introduction

This document provides a framework and best practices to evaluate image compression algorithms. This document provides a selection of evaluation tools that allow testing multiple features, including objective metric image quality, subjective metric image quality and codec algorithmic complexity. Which features of codecs should be tested and pass-fail criteria is beyond the scope of this document.

Information technology — Advanced image coding and evaluation —

Part 1: Guidelines for image coding system evaluation

1 Scope

This document recommends best practices for coding system evaluation of images and image sequences. This document defines a common vocabulary of terms for coding system evaluation and divides evaluation methods into three broad categories:

a) subjective assessment;

b) objective assessment;

c) computational assessment.

In addition to these broad assessment categories, this document discusses special care that is given for coding unusual imagery, e.g. high dynamic range or high colour depth.

A fourth assessment category, hardware complexity, is often important for real-time or computationally complex applications; however, it is outside the scope of this document.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

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

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
channel
one logical component of an image

Note 1 to entry: A channel may be a direct representation of one component from the bitstream, or may be generated by the application of a palette to a component from the bitstream.

[SOURCE: ISO/IEC 15444-1:2016, 3.17 – modified to move part of definition into a Note to entry]

3.2
codec
coding system
system comprising a *compressor* (3.6), a *decompressor* (3.8) and the compressor's bitstream output is compatible with the decompressor's bitstream input

3.3
component
two-dimensional array of samples

Note 1 to entry: An image typically consists of several components, for instance, representing red, green, and blue.

[SOURCE: ISO/IEC 15444-1:2016, 3.26 – modified to move part of definition into a Note to entry]

3.4
component bit depth
number of bits of precision of colour channels (or components) of an unencoded image

3.5
component number
number of colour channels (or components) encoded in an image

3.6
compressor
portion of a coding system that has a pixel stream and may have control metadata as its input and a coded bitstream as its output

3.7
constant bit rate
mode where the number of encoded bits from a portion of an image represented by a fixed number of *pixels* (3.16) does not vary compared to the number of encoded bits in any other equally sized portion of the same image

3.8
decompressor
portion of a *codec (coding system)* (3.2) that has a coded bitstream as its input and a *pixel* (3.16) stream as its output

3.9
drift
net generational loss of image quality if the output of a lossy image compression/reconstruction cycle is recompressed again under the same conditions by the same *codec* (3.2)

3.10
expert observer
observer that has expertise in image artefacts that may be introduced by the system under test or who has designed or participated in the selection of test content for the system under test

3.11
generational quality loss
measure of *quality loss* (3.17) between a reference image and a reconstruction of the same image after repetitive generations of encoding and decoding

3.12
horizontal pixel resolution
horizontal extent of the image in image *pixels* (3.16) where the horizontal extent may depend on the channel

3.13
idempotent
codec (3.2) that operates losslessly on its own decompression output

3.14
non-expert observer
naïve observer
observer that has no expertise in the image artefacts that may be introduced by the system under test

3.15
objective assessment
computational algorithmic process leading to a numerical score for all or a portion of an image under test

3.16
pixel
smallest element that is capable of generating the full intended functionality, e.g. colour and grey scale, of the display

Note 1 to entry: In a multicolour display, the smallest addressable element capable of producing the full colour range or the smallest element that is capable of generating the full functionality of the display.

3.17
quality loss
measure of the difference between a reference image and an encoded and reconstructed representation of the same image

3.18
sample
one unit of a grey scale or colour where an unencoded image comprises a plurality of these units

3.19
sample precision
bit depth of a given data type encoding the image

3.20
sample type
type of numeric value that contains *sample* (3.18) values to a resolution specified by *sample precision* (3.19) where types can include unsigned integers, signed integers and floating point or fixed point samples

3.21
sub-sample
sample (3.18) where the number of samples in either the horizontal dimension or the vertical dimension is not equal to the horizontal or vertical image dimension, respectively

3.22
subjective assessment
algorithmic process where recorded observations from human subjects (observers) lead to a numerical score for all or a portion of an image under test

3.23
variable bit rate
mode where the number of encoded bits in a portion of an image represented by a fixed number of *pixels* (3.16) can be different from the number of encoded bits in any other equally sized portion of the same image

3.24
vertical pixel resolution
vertical extent of the image in *pixels* (3.16) and the vertical extent may depend on the channel for subsampled images

4 Abbreviated terms

bpp	bits per pixel
CIE	International Commission on Illumination
CIEDE2000	CIE colour difference formula

CIELAB	CIE – Lab colour space
CIE-XYZ	CIE – XYZ colour space
CR	compression ratio
CSF	contrast sensitivity function
CW-SSIM	complex wavelet structural similarity index
DDP	degree of data parallelism
HDR	high dynamic range
HDR-VDP	high dynamic range visual difference predictor
HVS	human visual system
JND	just noticeable difference
LDR	low dynamic range, synonymous with SDR
MOS	mean opinion score
MSE	mean squared error
MSSIM	mean structural similarity index
MS-SSIM	multi scale structural similarity index
PSNR	peak signal-to-noise ratio
RDP	ratio of pixels to data parallelism
S-CIELAB	spatial extension to CIEDE2000
SDR	standard dynamic range, synonymous with LDR
SIMD	single instruction, multiple data
SSIM	structural similarity index
VDM	visual discrimination model
VDP	visual differences predictor

5 Selection and characteristics of test images

5.1 Common image characteristics

Image selection relies on a common vocabulary for describing image characteristics. This clause defines this vocabulary and the applicability to testing both standard and high dynamic range images.

For example, integer samples in range [0..1023] are here described as ten bit data, regardless of whether the samples are stored in 16 bit values or packed into ten bits each. Integer values in the range [-128..127] are here classified as 8 bit signed data because the data representation consists of one sign bit and seven magnitude bits.

The image dimension data consists of the full set of data defined above, that is, the number of channels, the width and height of each image channel, the sample type of each channel and the sample precision of each channel.

5.2 Bits per pixel

Bits per pixel (bpp) describes the compression performance of image compression codecs independent of the original image's sample size.

bpp, given in Formula (1), is defined independently of the image sample precision as the size of the compressed image stream L and the image dimensions, w and h:

$$\text{bpp} = \frac{8 \cdot L}{w \cdot h} \tag{1}$$

where

L is the compressed image stream, in bytes;

w is the width;

h is the height.

5.3 Compression ratio

Compression ratio (CR), given in Formula (2), describes the compression performance of image coding system dependent of the original image's sample size[6]:

$$\text{CR} = \frac{\sum_{c=0}^{d-1} b(c) \cdot w(c) \cdot h(c)}{8 \cdot L} \tag{2}$$

where

d is the number of channels of the image;

$w(c)$ is the horizontal extent of channel c;

$h(c)$ is the vertical extent of channel c;

$b(c)$ is the number of bits of sample precision in the samples of channel c.

5.4 Variation in bit rates

5.4.1 Constant bit rate systems

Constant bit rate systems have a constant pixels per unit of time input that matches the constant pixels per unit of time output without variation within an image. A test can verify if any bit rate variation is present. This restriction may not apply between two or more images.

5.4.2 Variable bit rate systems

For some applications, it is important that a coding system is able to generate a continuous stream of symbols, ensuring that some output is generated at least in every given time span, i.e. that the output bit rate does not vary too much over time. For example, carry-over resolution in arithmetic coding might cause arbitrary long delays in the output until the carry can be resolved.

For the purpose of this test, the output bit rate is defined as the number of output symbols generated for each input symbol, measured in dependence of the percentage of the input stream fed into the codec.

A measurement procedure to measure bit rate variations appears in Annex D.

5.5 Error resilience

In modern systems, error resiliency can be assisted by error markers in the bitstream or error resiliency can be part of transport layer capabilities. A coding system evaluation needs to take into consideration whether error resiliency is in a bitstream and if so, whether optional or intertwined and inseparable.

The best practices at the time of this document separates error resiliency by computing the efficiency of the algorithm to code images while assuming a perfect transmission medium. The ability to recover errors can be added either through resiliency markers, forward error correction or merely parity checking to identify but not correct errors.

If separable, the topic is outside the scope of this document and codec testing should assume no error introduction in the bitstream.

If error markers and error handling markers are not separable from the coded bitstream, the coding system efficiency will include such markers.

5.6 Recursive compression assessment

Generation loss is a loss in image quality if the output of a lossy image compression/decompression cycle is recompressed again under the same conditions by the same compression/decompression. If this recompression is repeated over several cycles, this can result in severe degradation of image quality[26].

Generation loss limits the number of repeated compressions/decompressions in an image processing chain if repeated recompression generates severely more distortion than a single compression/decompression cycle. This subclause distinguishes between drift and quality loss. While the former is due to a systematic DC error often due to mis-calibration in the colour transformation or quantization, the latter covers all other error sources, as well as, for example, due to limited precision in the image transformation implementation.

A measurement procedure to measure generational quality loss appears in Annex D.

5.7 Image selection

Colour content and categories of images to consider when testing a codec include continuous tone images, black and white or half tones. Test material should reflect the potential applications in which a coding system will be used. The following examples represent common image categories for evaluation:

a) natural scenes;

b) portraits with differing skin tones;

c) compound (multi-layer);

d) photo-realistic synthetic;

e) graphics and animations;

f) text and web pages;

g) engineered test patterns.

If the coding system is intended for specific image types or applications, such as medical imaging, a set of images appropriate to the application should be the test set.

Image size used during testing should be appropriate for the application, not very much smaller or larger than targeted in typical usage.

6 Best practices of subjective image quality assessments

6.1 Goals of subjective assessment

Some subjective image assessment methods are likely to reflect the human notion of quality by anticipating the reactions of those who might view the tested systems. While other subjective image assessment methods can determine if some artefacts are visually discernible and likely to adversely affect image quality. These methods become the best quality assessment methods. However, they are very time demanding and they might eventually become very expensive, because of the cost of the viewers and also of the system under test implementation.

Test evaluations can be application specific, for example, according to Rec. ITU-R BT.500[7]:

> "*Subjective assessment methods are used to establish the performance of television systems using measurements that more directly anticipate the reactions of those who might view the systems tested. In this regard, it is understood that it may not be possible to fully characterize system performance by objective means; consequently, it is necessary to supplement objective measurements with subjective measurements.*"

This document suggests that best practice should separate applications from the image quality evaluation to the best extent possible. Subjective assessment methodology recommended herein follows this guideline.

Best practices in this document draw from the psychophysical experimental method standardized in ISO 29462-2 for photography[3] and extended the methods for electronic displays.

Some applications will have specific goals differing from general practice, such as, radiological images[8].

6.2 Subjective assessment evaluation procedures

6.2.1 Observer selection

Evaluators should prefer naïve observers for most general viewing or entertainment applications. In the case of specialized imaging, such as, medical or structural engineering, an expert observer who can discern defects from artefacts is needed.

6.2.2 Visual acuity

Common to all subjective evaluation procedures, observers will need to demonstrate meet a well-defined visual acuity. Sometimes colour vision is not tested.

The following recommendations usually apply.

a) Test for visual acuity with or without corrective lens, either glasses or contacts that do not have multiple focal lengths, e.g. progressive, bifocal or trifocal corrective lens.

b) Verify normal visual acuity by using a Snellen or Landolt test charts where the observer reads at 20/20 from 50 cm.

c) If screening for normal colour vision, verify by testing with Ishihara plates or equivalent.

Evaluators may refer to ISO/IEC 29170-2 for examples of tools that help assess an observer's visual acuity[5].

6.2.3 Number of observers

The number of observers is dependent on the evaluation system. For example, according to Rec. ITU-R BT.500: