FINAL
DRAFT

INTERNATIONAL
STANDARD

**ISO/IEC
FDIS
18477-3**

ISO/IEC JTC **1**/SC **29**

Secretariat: **JISC**

Voting begins on:
2023-08-23

Voting terminates on:
2023-10-18

Information technology — Scalable compression and coding of continuous-tone still images —

Part 3:
Box file format

Technologies de l'information — Compression échelonnable et codage d'images plates en ton continu —

Partie 3: Format de la liste de fichiers

Reference number
ISO/IEC FDIS 18477-3:2023(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

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.

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 or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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

For an explanation of 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 www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

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*.

This second edition cancels and replaces the first edition (ISO/IEC 18477-3:2015), which has been technically revised.

The main changes are as follows:

— editorial improvements on the usage of the JPEG XT marker segment.

A list of all parts in the ISO/IEC 18477 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

This document is an extension of ISO/IEC 18477-1, a compression system for continuous-tone digital still images, which is backwards compatible with Rec. ITU-T T.81 | ISO/IEC 10918-1. This means that legacy applications conforming to Rec. ITU-T T.81 | ISO/IEC 10918-1 will be able to reconstruct streams generated by an encoder conforming to this document, although it is possible that they will not be able to reconstruct such streams in full dynamic range or quality or using other features defined in this document.

This document provides a flexible and extensible framework to enrich ISO/IEC 18477-1 conforming codestreams with side-channels and metadata. The syntax chosen in this document defines a mechanism to embed syntax elements denoted as “boxes” into Rec. ITU-T T.81 | ISO/IEC 10918-1 conforming codestreams. The box syntax used in this document is identical to that defined in the JPEG series, for example JPEG 2000 image coding system (Rec. ITU-T T.800 | ISO/IEC 15444-1). Boxes will then carry either additional image data, to enable encoding of images of higher bit depth, high-dynamic range (HDR), including alpha channels, etc., or metadata that describes the decoding process of the legacy Rec. ITU-T T.81 | ISO/IEC 10918-1 codestream and the side channels to an extended or HDR image.

This document specifies an extensible file format, denoted as JPEG XT, which is built on top of the existing Rec. ITU-T T.81 | ISO/IEC 10918-1 codestream definition. While typically, file formats encapsulate codestreams by means of additional syntax elements such as boxes, the file format structure specified in this document embeds the syntax elements of the file format, called boxes, into the codestream. The necessity for this unusual arrangement is the backwards compatibility to the legacy standard and the application toolchain built around it. This means that legacy applications conforming to Rec. ITU-T T.81 | ISO/IEC 10918-1 will be able to decode image information embedded in files conforming to this document, although they will only be able to recover a three component, 8 bit per sample, lower quality version of the image described by the full file.

For more demanding applications, it is not uncommon to use a bit depth of 16, providing 65 536 representable values to describe each channel within a pixel, resulting on over 2.8×10^{14} representable colour values. In some less common scenarios, even greater bit depths are used, and sometimes the dynamic range of the image is so high that a floating-point based encoding is desirable. In addition to image information, some applications also require an additional opacity channel, a feature not available from the legacy standard.

Most common photo and image formats use an 8-bit or 16-bit unsigned integer value to represent some function of the intensity of each colour channel. While it can be theoretically possible to agree on one method for assigning specific numerical values to real world colours, doing so is not practical. Since any specific device has its own limited range for colour reproduction, the device’s range can be a small portion of the agreed-upon universal colour range. As a result, such an approach is an extremely inefficient use of the available numerical values, especially when using only 8 bits (or 256 unique values) per channel. To represent pixel values as efficiently as possible, devices use a numeric encoding optimized for their own range of possible colours or gamut.

JPEG XT is designed to extend the legacy JPEG standard towards higher bit depth, higher dynamic range, and wide colour gamut content, while simultaneously allowing legacy applications to decode the image data in the codestream to a standard low-dynamic range (LDR) image represented by only 8 bits per channel. The goal is to provide a backwards compatible coding specification that allows legacy applications and existing toolchains to continue to operate on codestreams conforming to this document.

JPEG XT has been designed to be backwards compatible to legacy applications while at the same time having a small coding complexity. JPEG XT uses, whenever possible, functional blocks of Rec. ITU-T T.81 | ISO/IEC 10918-1 to extend the functionality of the legacy JPEG coding system.

Information technology — Scalable compression and coding of continuous-tone still images —

Part 3:
Box file format

1 Scope

This document specifies box-based container format, referred to as JPEG XT, which is designed primarily for continuous-tone photographic content.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements 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/IEC 18477-1, *Information technology — Scalable compression and coding of continuous-tone still images — Part 1: Core coding system specification*

Rec. ITU-T T.81 | ISO/IEC 10918-1, Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines

Rec. ITU-T T.871 | ISO/IEC 10918-5, Information technology — Digital compression and coding of continuous-tone still images: JPEG File Interchange Format (JFIF) — Part 5:

3 Terms, definitions, abbreviated terms and symbols

3.1 Terms and definitions

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

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

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

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

3.1.1
ASCII encoding
encoding of text characters and text strings according to ISO/IEC 10646-1

3.1.2
base decoding path
process of decoding legacy codestream and refinement data to the base image, jointly with all further steps, until residual data is added to the values obtained from the residual codestream

3.1.3
base image
collection of sample values obtained by entropy, decoding the discrete cosine transformation (DCT) coefficients of the legacy codestream and the refinement codestream, and inversely DCT transforming them jointly

3.1.4
bit stream
partially encoded or decoded sequence of bits comprising an entropy-coded segment

3.1.5
box
structured collection of data describing the image or the image-decoding process embedded into one or multiple APP_{11} marker segments

Note 1 to entry: See Annex A for the definition of boxes.

3.1.6
byte
group of 8 bits

3.1.7
compression
reduction in the number of bits used to represent source image data

3.1.8
component
two-dimensional array of samples having the same designation in the output or display device

Note 1 to entry: An image typically consists of several components, e.g. red, green and blue.

3.1.9
continuous-tone image
image whose components have more than one bit per sample

3.1.10
decoder
embodiment of a decoding process

3.1.11
decoding process
process which takes as its input compressed image data and outputs a continuous-tone image

3.1.12
encoder
embodiment of an encoding process

3.1.13
encoding process
process which takes as its input a continuous-tone image and outputs compressed image data

3.1.14
extension image
sample values as reconstructed by inverse quantization and inverse discrete cosine transformation (DCT) applied to the entropy-decoded coefficients described by the refinement scan, residual scan and residual refinement scans

3.1.15
high-dynamic range
HDR
image or image data comprised of more than 8 bits per sample

3.1.16
legacy codestream
collection of markers and syntax elements defined by Rec. ITU-T T.81 | ISO/IEC 10918-1 without any syntax elements defined by ISO/IEC 18477-1, ISO/IEC 18477-2, ISO/IEC 18477-3

Note 1 to entry: In this definition, the legacy codestream consists of the collection of all markers except those APP_{11} markers that describe JPEG XT boxes by the syntax defined in Annex A.

3.1.17
legacy decoder
embodiment of a decoding process conforming to Rec. ITU-T T.81 | ISO/IEC 10918-1, confined to the lossy discrete cosine transformation (DCT) process and the baseline, sequential or progressive modes, decoding at most four components to 8 bits per component

3.1.18
lossless
encoding and decoding processes and procedures in which the output of the decoding procedure(s) is identical to the input to the encoding procedure(s)

3.1.19
lossy
encoding and decoding processes which are not lossless

3.1.20
low-dynamic range
LDR
image or image data comprised of data with no more than 8 bits per sample

3.1.21
marker
2-byte code in which the first byte is hexadecimal FF and the second byte is a value between 1 and hexadecimal FE

3.1.22
marker segment
marker together with its associated set of parameters

3.1.23
pixel
collection of sample values in the spatial image domain having all the same sample coordinates

Note 1 to entry: A pixel may consist of three samples describing its red, green and blue value.

3.1.24
point transformation
scaling of a sample or discrete cosine transformation (DCT) coefficient by a factor

3.1.25
precision
number of bits allocated to a particular sample or discrete cosine transformation (DCT) coefficient

3.1.26
procedure
set of steps which accomplishes one of the tasks which comprises an encoding or decoding process

3.1.27
residual decoding path
collection of operations applied to the entropy coded data contained in the residual data box and residual refinement scan boxes up to the point where this data is merged with the base image to form the final output image

3.1.28
residual image
sample values as reconstructed by inverse quantization and inverse discrete cosine transformation (DCT) applied to the entropy-decoded coefficients described by the residual scan and residual refinement scans

3.1.29
refinement scan
additional pass over the image data that is invisible to legacy decoders, which provides additional least significant bits to extend the precision of the discrete cosine transformation (DCT) transformed coefficients

Note 1 to entry: Refinement scans can be either applied in the base or residual decoding path.

3.1.30
sample
one element in the two-dimensional image array which comprises a component

3.1.31
sample grid
common coordinate system for all samples of an image such that the samples at the top left edge of the image have the coordinates (0, 0), the first coordinate increases towards the right, the second towards the bottom

3.1.32
superbox
box that carries other boxes as payload data

3.1.33
zero byte
0x00 byte

3.2 Abbreviated terms

ASCII American Standard Code for Information Interchange

LSB least significant bit

MSB most significant bit

HDR high-dynamic range

IDR intermediate-dynamic range

JPEG informal name of the committee that created this document

LDR low-dynamic range

TMO tone mapping operator

DCT discrete cosine transformation

3.3 Symbols

X Width of the sample grid in positions

Y Height of the sample grid in positions

Nf Number of components in an image

$s_{i,x}$ Subsampling factor of component i in horizontal direction

$s_{i,y}$ Subsampling factor of component i in vertical direction

H_i Subsampling indicator of component i in the frame header

V_i Subsampling indicator of component i in the frame header

$v_{x,y}$ Sample value at the sample grid position x, y

R_h Additional number of DCT coefficient bits represented by refinement scans in the base decoding path. $8+R_h$ is the number of non-fractional bits (i.e. bits in front of the "binary dot") of the output of the inverse DCT process in the base decoding path.

R_r Additional number of DCT coefficient bits represented by refinement scans in the residual decoding path. $P+R_r$ is the number of non-fractional bits of the output of the inverse DCT process in the residual decoding path, where P is the frame-precision of the residual image as recorded in the frame header of the residual codestream.

R_b Additional bits in the HDR image. $8+R_b$ is the sample precision of the reconstructed HDR image.

4 Conventions

4.1 Conformance language

The keyword "reserved" indicates a provision that is not specified at this time, shall not be used, and may be specified in the future. The keyword "forbidden" indicates "reserved" and, in addition, indicates that the provision will never be specified in the future.

4.2 Operators

NOTE Many of the operators used in this document are similar to those used in the C programming language.

4.2.1 Arithmetic operators

+	Addition
−	Subtraction (as a binary operator) or negation (as a unary prefix operator)
×	Multiplication
/	Division without truncation or rounding
umod	x umod a is the unique value y between 0 and a-1 for which y+Na = x with a suitable integer N

4.2.2 Logical operators

||	Logical OR
&&	Logical AND
!	Logical NOT
∈	x ∈ {A, B} is defined as (x == A || x == B)
∉	x ∉ {A, B} is defined as (x != A && x != B)

4.2.3 Relational operators

>	Greater than
>=	Greater than or equal to
<	Less than
<=	Less than or equal to
==	Equal to
!=	Not equal to

4.2.4 Precedence order of operators

Operators are listed below in descending order of precedence. If several operators appear in the same line, they have equal precedence. When several operators of equal precedence appear at the same level in an expression, evaluation proceeds according to the associativity of the operator either from right to left or from left to right.

Operators	Type of operation	Associativity
(), [ ], .	Expression	Left to Right
–	Unary negation	
×, /	Multiplication	Left to Right
umod	Modulo (remainder)	Left to Right
+, –	Addition and Subtraction	Left to Right
< , >, <=, >=	Relational	Left to Right

4.2.5 Mathematical functions

$\lceil x \rceil$	Ceiling of *x*. Returns the smallest integer that is greater than or equal to *x*.
$\lfloor x \rfloor$	Floor of *x*. Returns the largest integer that is lesser than or equal to *x*.
$\lvert x \rvert$	Absolute value, is –x for x < 0, otherwise *x*.
sign(*x*)	Sign of *x*, 0 if *x* is 0, +1 if *x* is positive, –1 if *x* is negative.

clamp(x,min,max)	Clamps x to the range [min, max]: returns min if $x < min$, max if $x > max$ or otherwise x.
x^a	Raises the value of x to the power of a. x is a non-negative real number, a is a real number. x^a is equal to $\exp(a \times \log(x))$ where exp is the exponential function and log() the natural logarithm. If x is 0 and a is positive, x^a is defined to be 0.

5 Overview

5.1 General

This clause gives an informative overview of the elements specified in this document. It also introduces many of the terms which are defined in Clause 3.

There are three elements specified in this document:

a) An "encoder" is an embodiment of an "encoding process". An encoder takes as input "digital source image data" and "encoder specifications", and by means of a specified set of "procedures" generates as output a ″codestream".

b) A "decoder" is an embodiment of a "decoding process". A decoder takes as input a codestream, and by means of a specified set of procedures generates as output "digital reconstructed image data".

c) The "codestream" is a compressed image data representation which includes all necessary data to allow a (full or approximate) reconstruction of the sample values of a digital image. Additional data can be required that define the interpretation of the sample data, such as colour space or the spatial dimensions of the samples.

5.2 High-level overview on JPEG XT

The high-level syntax of an ISO/IEC 18477-3 conforming codestream is identical to that defined in ISO/IEC 18477-1, which is a subset of the syntax defined in Rec. ITU-T T.81 | ISO/IEC 10918-1. Marker definitions and the syntax of the markers defined in Rec. ITU-T T.81 | ISO/IEC 10918-1 remain in force and unchanged. However, this document defines the APP_{11} marker, reserved in the legacy Recommendation | Standard for encoding additional syntax elements. Legacy decoders will skip and ignore such marker elements, and hence will only be able to decode the image encoded by the legacy syntax elements. This part of a JPEG XT file will be denoted the legacy codestream in the following.

This document extends the legacy standard by a syntax element called “box”, using the APP_{11} marker to hide the extended syntax elements from legacy applications. Boxes and their encoding are specified in Annex A. A common set of boxes used by ISO/IEC 18477-6, ISO/IEC 18477-7, ISO/IEC 18477-8 and ISO/IEC 18477-9 are defined in Annex B. A box can either include additional metadata required to decode the complete codestream to full precision, full dynamic range or without loss, or can contain entropy coded image data itself.

How entropy coded data from the side-channels contained in the boxes and entropy coded data in the legacy codestream are merged together is application dependent and defined in ISO/IEC 18477-6, ISO/IEC 18477-7, ISO/IEC 18477-8 or ISO/IEC 18477-9. It is beyond the scope of this document to define this process.

5.3 Encoder requirements

An encoder is only required to meet the compliance tests and to generate the codestream according to the syntax defined in this document. How the codestream is algorithmically constructed and how the boxes are laid out is implementation-specific and not within scope of this document. ISO/IEC 18477-6, ISO/IEC 18477-7, ISO/IEC 18477-8 and ISO/IEC 18477-9 can, however, define additional restrictions and requirements, either within the standard itself, or within profiles that restrict the freedom of the encoder further.

An encoder claiming to conform to one of these profiles then shall conform to the syntax constraints defined in the corresponding profile of the corresponding part of ISO/IEC 18477-6, ISO/IEC 18477-7, ISO/IEC 18477-8 or ISO/IEC 18477-9.

5.4 Decoder requirements

A decoding process converts compressed image data to reconstructed image data. A decoder shall interpret the syntax of the box structures, namely the packaging of boxes correctly into APP markers specified in Annex A. It is not required, however, for a conforming decoder to be capable of interpreting the semantics of all box types defined in this document. A decoder implementation should skip over boxes it is unable or not willing to support unless such a box is indicated as a mandatory box in the profile and part of the ISO/IEC 18477 series to which the decoder claims to conform.