INTERNATIONAL STANDARD

ISO/IEC 18477-3

First edition

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

PROOF/ÉPREUVE

Reference number
ISO/IEC 18477-3:2015(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

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 meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ISO/IEC 18477 contains the following parts under the general title *Information technology — Scalable compression and coding of continuous-tone still images:*

— *Part 1: Scalable compression and coding of continuous-tone still images*

— *Part 2: Extensions for high dynamic range images*

— *Part 3: Box file format*

— *Part 6: IDR Integer Coding*

— *Part 7: HDR Floating-Point Coding*

— *Part 8: Lossless and Near-lossless Coding*

— *Part 9: Alpha Channel Coding*

The following parts are under preparation:

— *Part 4: Conformance testing*

— *Part 5: Reference software*

Introduction

This part of ISO/IEC 18477 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 here rather 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; that is, 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 part of ISO/IEC 18477, though will only be able to recover a three component, 8 bits 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 $2{,}8 \times 10^{14}$ 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 might 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 may 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 bitdepth, higher dynamic range, wide colour gamut content while simultaneously allowing legacy applications to decode the image data in the codestream to a standard low dynamic range image represented by only eight 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 part of ISO/IEC 18477.

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.

This part of ISO/IEC 18477 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. That is, 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 part of ISO/IEC 18477, though will possibly not be able to reconstruct such streams in full dynamic range, full quality or other features defined in this part of ISO/IEC 18477.

The aim of this part of ISO/IEC 18477 is to provide a flexible and extensible framework to enrich ISO/IEC 18477-1 compliant codestreams with side-channels and metadata. The syntax chosen in this part of ISO/IEC 18477 defines a mechanism to embed syntax elements denoted as “Boxes” into Rec. ITU-T T.81 | ISO/IEC 10918-1 compliant codestreams. The box syntax used here is identical to that defined in the JPEG family of standards, for example JPEG 2000 (Rec. ITU-T T. 800 | ISO/IEC 15444-1). Boxes will then carry either additional image data, to enable encoding of images of higher bitdepth, high dynamic range, include alpha channels etc., or will carry 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 high dynamic range image.

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

Part 3: **Box file format**

1 Scope

This part of ISO/IEC 18477 specifies a coding format, referred to as JPEG XT, which is designed primarily for continuous-tone photographic content.

2 Normative references

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

ISO/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*

ISO/IEC 18477-1:2015, *Information technology — Scalable compression and coding of continuous-tone still images — Part 1: Scalable compression and coding of continuous-tone still images*

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*

3 Terms, definitions, abbreviated terms and symbols

3.1 Terms and definitions

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

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

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 DCT coefficients of the legacy codestream and the refinement codestream, and inversely DCT transforming them jointly

3.1.4
binary decision
choice between two alternatives

3.1.5
bit stream
partially encoded or decoded sequence of bits comprising an entropy-coded segment

3.1.6
block
8 × 8 array of samples or an 8 × 8 array of DCT coefficient values of one component

3.1.7
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

3.1.8
byte
group of 8 bits

3.1.9
coder
embodiment of a coding process

3.1.10
coding
encoding or decoding

3.1.11
coding model
procedure used to convert input data into symbols to be coded

3.1.12
(coding) process
general term for referring to an encoding process, a decoding process, or both

3.1.13
compression
reduction in the number of bits used to represent source image data

3.1.14
component
two-dimensional array of samples having the same designation in the output or display device. An image typically consists of several components, e. g. red, green and blue

3.1.15
continuous tone image
image whose components have more than one bit per sample

3.1.16
decoder
embodiment of a decoding process

3.1.17
decoding process
process which takes as its input compressed image data and outputs a continuous-tone image

3.1.18
encoder
embodiment of an encoding process

3.1.19
encoding process
process which takes as its input a continuous-tone image and outputs compressed image data

3.1.20
entropy decoder
embodiment of an entropy decoding procedure

3.1.21
entropy decoding
lossless procedure which recovers the sequence of symbols from the sequence of bits produced by the entropy encoder

3.1.22
entropy encoder
embodiment of an entropy encoding procedure

3.1.23
entropy encoding
lossless procedure which converts a sequence of input symbols into a sequence of bits such that the average number of bits per symbol approaches the entropy of the input symbols

3.1.24
high dynamic range
image or image data comprised of more than eight bits per sample

3.1.25
Intermediate dynamic range
image or image data comprised of more than eight bits per sample

3.1.26
Joint Photographic Experts Group
JPEG
informal name of the committee which created this part of ISO/IEC 18477

Note 1 to entry: The "joint" comes from the ITU and ISO/IEC collaboration.

3.1.27
legacy codestream
collection of markers and syntax elements defined by Rec. ITU-T T.81 | ISO/IEC 10918-1 bare any syntax elements defined by the family ISO/IEC 18477 standards

Note 1 to entry: That is, 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.

3.1.28
legacy decoding path
collection of operations to be performed on the entropy coded data as described by Rec. ITU-T T.81 | ISO/IEC 10918-1 jointly with the Legacy Refinement scans before this data is merged with the residual data to form the final output image

3.1.29
legacy decoder
embodiment of a decoding process conforming to ITU. T Rec. T.81 | ISO/IEC 10918-1, confined to the lossy DCT process and the baseline, sequential or progressive modes, decoding at most four components to eight bits per component

3.1.30
legacy image
arrangement of sample values as described by applying the decoding process described by Rec. ITU-T T.81 | ISO/IEC 10918-1 on the entropy coded data as defined by said standard

3.1.31
lossless
descriptive term for 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)

3.1.32
lossless coding
mode of operation which refers to any one of the coding processes defined in this part of ISO/IEC 18477 in which all of the procedures are lossless

3.1.33
lossy
descriptive term for encoding and decoding processes which are not lossless

3.1.34
low dynamic range
image or image data comprised of data with no more than eight bits per sample

3.1.35
marker
two-byte code in which the first byte is hexadecimal FF and the second byte is a value between 1 and hexadecimal FE

3.1.36
marker segment
marker together with its associated set of parameters

3.1.37
pixel
collection of sample values in the spatial image domain having all the same sample coordinates, e. g. a pixel may consist of three samples describing its red, green and blue value

3.1.38
point transform
scaling of a sample or DCT coefficient by a factor

3.1.39
precision
number of bits allocated to a particular sample or DCT coefficient

3.1.40
procedure
set of steps which accomplishes one of the tasks which comprise an encoding or decoding process

3.1.41
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

3.1.42
residual image
extension image
sample values as reconstructed by inverse quantization and inverse DCT transformation applied to the entropy-decoded coefficients described by the residual scan and residual refinement scans

3.1.43
residual scan
additional pass over the image data invisible to legacy decoders which provides additive and/or multiplicative correction data of the base scans to allow reproduction of high dynamic range or wide colour gamut data

3.1.44
refinement scan
additional pass over the image data invisible to legacy decoders which provides additional least significant bits to extend the precision of the DCT transformed coefficients. Refinement scans can be either applied in the base or residual decoding path

3.1.45
sample
one element in the two-dimensional image array which comprises a component

3.1.46
sample grid
common coordinate system for all samples of an image

Note 1 to entry: 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.

3.1.47
superbox
box that carries other boxes as payload data

3.1.48
zero byte
0x00 byte

3.1.49
zig-zag sequence
specific sequential ordering of the DCT coefficients from (approximately) lowest spatial frequency to highest

3.2 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 invers 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+Rb is the sample precision of the reconstructed HDR image.

3.3 Abbreviated terms

For the purposes of this part of ISO/IEC 18477, the following abbreviations apply.

ASCII American Standard Code for Information Interchange

LSB Least Significant Bit

MSB Most Significant Bit

HDR High Dynamic Range

IDR Intermediate Dynamic Range

LDR Low Dynamic Range

TMO Tone Mapping Operator

DCT Discrete Cosine Transformation

4 Conventions

4.1 Conformance language

This part of ISO/IEC 18477 consists of normative and informative text.

Normative text is that text which expresses mandatory requirements. The word “shall” is used to express mandatory requirements strictly to be followed in order to conform to this part of ISO/IEC 18477 and from which no deviation is permitted. A conforming implementation is one that fulfils all mandatory requirements.

Informative text is that is potentially helpful to the user, but not indispensable and can be removed, changed or added editorially without affecting interoperability. All text in this part of ISO/IEC 18477 is normative, with the following exceptions: the Introduction, any parts of the text that are explicitly labelled as “informative”, and statements appearing with the preamble “NOTE” and behaviour described using the word “should”. The word “should” is used to describe behaviour that is encouraged but is not required for conformance to this part of ISO/IEC 18477.

The keywords “may” and “need not” indicate a course of action that is permissible in a conforming implementation.

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 part of ISO/IEC 18477 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

$\lceil x \rceil$	Ceil 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.