



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Series T

Supplement 2
(03/2011)

SERIES T: TERMINALS FOR TELEMATIC SERVICES

JPEG XR image coding system – System architecture

CAUTION !

PREPUBLISHED RECOMMENDATION

This prepublication is an unedited version of a recently approved Recommendation. It will be replaced by the published version after editing. Therefore, there will be differences between this prepublication and the published version.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this publication, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this publication is voluntary. However, the publication may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the publication is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the publication is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this publication may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the publication development process.

As of the date of approval of this publication, ITU [had/had not] received notice of intellectual property, protected by patents, which may be required to implement this publication. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2011

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

JPEG XR image coding system – System architecture

Summary

This Supplement is technically aligned with ISO/IEC 29199-1 but is not published as identical text. It was drafted in collaboration with ISO/IEC JTC 1/SC 29/WG 1 (which is informally known as "JPEG").

This document provides a technical overview and informative guidelines for applications of the JPEG XR image coding system as normatively specified in ITU-T Rec. T.832 | ISO/IEC 29199-2, ITU-T Rec. T.833 | ISO/IEC 29199-3, ITU-T Rec. T.834 | ISO/IEC 29199-4, and ITU-T Rec. T.835 | ISO/IEC 29199-5. The overview of JPEG XR coding technology includes a description of the supported image formats, the internal data processing hierarchy and data structures, the image tiling design supporting hard and soft tiling of images, the lapped bi-orthogonal transform, supported quantization modes, adaptive coding and scanning of coefficients, entropy coding, and finally the codestream structure. This overview provides a basic understanding of how a JPEG XR encoder works and the various modes it supports. It also compares the JPEG XR design with those of baseline JPEG (Rec. ITU-T T.81 | ISO/IEC 10918-1) and JPEG 2000 (Rec. ITU-T T.800 | ISO/IEC 15444-1). Following the overview is a discussion of the use of JPEG XR for High Dynamic Range (HDR) image coding. The next section reviews various JPEG XR profiles and describes their target applications. An encoding practices section provides general encoding guidelines and guidelines for encoding for providing random access functionality, including an analysis of tile size selection trade-offs. A decoding process functionality section describes the decoding process and output colour conversions, and describes how to make use of JPEG XR scalability features in a decoding application. These scalability features include resolution, quality and spatial random access scalabilities. Finally, the last section of report describes codestream manipulations in the compressed domain. This section describes methods for trimming a codestream to extract a smaller codestream, switching between spatial and frequency codestream modes, rotation and flipping of images, extraction of a region of interest in the compressed domain, switching between interleaved and planar alpha planes, and modifying the tile structure of an image. The technical report is intended to help application developers to understand the JPEG XR design and to provide assistance in making effective use of its capabilities.

CONTENTS

Page

Foreword	4
1 Scope	5
2 Terms and definitions	5
3 Abbreviations	8
4 The JPEG XR image coding system	8
5 General overview of technical design	9
5.1 Basic technology structure	9
5.2 Supported image format types	10
5.3 Decoded image structure and interpretation	10
5.4 Data processing hierarchy and structures	11
5.5 The JPEG XR transform structure and hierarchy	12
5.6 Handling of image and tile boundaries	13
5.7 Quantization and lossless representation	14
5.7.1 Overall quantization design concepts	14
5.7.2 Quantization control on a spatial region basis	14
5.7.3 Quantization control on a frequency band basis	14
5.7.4 Quantization control on a colour plane component basis	15
5.7.5 Quantization control type combinations	15
5.8 Prediction of transform coefficients and coded block patterns	15
5.9 Adaptive ordering of coefficient scanning pattern	15
5.10 Entropy coding of transform coefficients	15
5.11 Codestream structure	16
6 JPEG XR design in relation to baseline JPEG and JPEG 2000	17
6.1 General	17
6.2 Image area partitions	18
6.3 Image fidelity refinement	18
7 High dynamic range (HDR) image coding	18
7.1 HDR formats supported in JPEG XR	18
7.2 HDR signal processing design in JPEG XR	19
7.3 Examples of HDR applications for JPEG XR	19
8 JPEG XR profiles and levels	19
8.1 Overview of profiles and levels	19
8.2 Sub-Baseline profile	20
8.3 Baseline profile	20
8.4 Main profile	20
8.5 Advanced profile	20
8.6 Levels	20
9 JPEG XR encoding practices	20
9.1 General encoding guidelines	20
9.2 Encoding for random access	21
9.3 Guidelines for tile size selection	21
10 The JPEG XR decoding process functionality	23
10.1 JPEG XR decoding process structure	23
10.2 Output colour conversion	24
10.3 Resolution scalability at decoder	24
10.3.1 General	24
10.3.2 DC-only image decoding	24
10.3.3 DC plus LP image decoding	24
10.4 Quality scalability at decoder	24
10.5 Spatial random access at decoder	25

11	JPEG XR codestream compressed-domain manipulation	25
11.1	General	25
11.2	Flexbits trimming	25
11.3	Flexbits and HP band elimination	25
11.4	Flexbits and HP and LP band elimination	25
11.5	Spatial versus frequency codestream mode switching	25
11.6	Rotation and flip	26
11.7	Compressed-domain region of interest extraction	26
11.8	Switching between interleaved and planar alpha planes	26
11.9	Compressed-domain retiling	26

Foreword

This document provides a technical overview and informative guidelines for applications of the JPEG XR image coding system as normatively specified in ITU-T Rec. T.832 | ISO/IEC 29199-2, ITU-T Rec. T.833 | ISO/IEC 29199-3, ITU-T Rec. T.834 | ISO/IEC 29199-4, and ITU-T Rec. T.835 | ISO/IEC 29199-5.

Copyright International Telecommunication Union 2011. All rights reserved.

INFORMATION TECHNOLOGY – JPEG XR IMAGE CODING SYSTEM – SYSTEM ARCHITECTURE

1 Scope

This document provides a technical overview and informative guidelines for applications of the JPEG XR image coding system as normatively specified in ITU-T Rec. T.832 | ISO/IEC 29199-2, ITU-T Rec. T.833 | ISO/IEC 29199-3, ITU-T Rec. T.834 | ISO/IEC 29199-4, and ITU-T Rec. T.835 | ISO/IEC 29199-5. The overview of JPEG XR coding technology includes a description of the supported image formats, the internal data processing hierarchy and data structures, the image tiling design supporting hard and soft tiling of images, the lapped bi-orthogonal transform, supported quantization modes, adaptive coding and scanning of coefficients, entropy coding, and finally the codestream structure. This overview provides a basic understanding of how a JPEG XR encoder works and the various modes it supports. It also compares the JPEG XR design with those of baseline JPEG (Rec. ITU-T T.81 | ISO/IEC 10918-1) and JPEG 2000 (Rec. ITU-T T.800 | ISO/IEC 15444-1). Following the overview is a discussion of the use of JPEG XR for High Dynamic Range (HDR) image coding. The next section reviews various JPEG XR profiles and describes their target applications. An encoding practices section provides general encoding guidelines and guidelines for encoding for providing random access functionality, including an analysis of tile size selection trade-offs. A decoding process functionality section describes the decoding process and output colour conversions, and describes how to make use of JPEG XR scalability features in a decoding application. These scalability features include resolution, quality and spatial random access scalabilities. Finally, the last section of report describes codestream manipulations in the compressed domain. This section describes methods for trimming a codestream to extract a smaller codestream, switching between spatial and frequency codestream modes, rotation and flipping of images, extraction of a region of interest in the compressed domain, switching between interleaved and planar alpha planes, and modifying the tile structure of an image. The technical report is intended to help application developers to understand the JPEG XR design and to provide assistance in making effective use of its capabilities.

2 Terms and definitions

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

- 2.1 adaptive coefficient normalization:** A parsing sub-process where *transform coefficients* are dynamically partitioned into a VLC-coded part and a fixed-length coded part, in a manner designed to control (i.e., "normalize") bits used to represent the VLC-coded part. The fixed-length coded part of *DC coefficients* and *low-pass coefficients* is called *FLC refinement* and the fixed-length coded part of *high-pass coefficients* is called *flexbits*.
- 2.2 adaptive inverse scanning:** A parsing sub-process where the *zigzag scan order* associated with a set of *transform coefficients* is dynamically modified, based on the statistics of previously-parsed *transform coefficients*.
- 2.3 adaptive VLC:** A parsing sub-process where the code table associated with VLC parsing of a particular *syntax element* is switched, among a finite set of fixed tables, based on the statistics of previously-parsed instances of this syntax element.
- 2.4 alpha image plane:** An optional secondary *image plane* associated with an *image*, of the same dimensions as the *luma component* of the *primary image plane*. The *alpha image plane* has one *component*, a *luma component*.
- 2.5 block:** An $m \times n$ array of *samples*, or an $m \times n$ array of *transform coefficients*.
- 2.6 block index:** An integer in the range 0 to 15, identifying, by its position in *raster scan order*, a particular 4×4 *block*, within a partition of a 16×16 *block* into 16 4×4 *blocks*.
- 2.7 byte:** A sequence of 8 bits.
- 2.8 chroma:** A *component* of the *primary image plane* with non-zero index, or the *transform coefficients* and sample values associated with this *component*.

- 2.9 codestream:** A sequence of bits contained in a sequence of *bytes* from which syntax elements are parsed, such that the most significant bit of the first *byte* is the first bit of the *codestream*, the next most significant bit of the first *byte* is the second bit of the *codestream*, and so on, to the least significant bit of the first *byte* (which is the eighth bit of the *codestream*), followed by the most significant bit of the second *byte* (which is the ninth bit of the *codestream*), and so on, up to and including the least significant bit of the last *byte* of the sequence of *bytes* (which is the last bit of the *codestream*).
- 2.10 component:** One of the arrays of samples associated with an *image plane*.
- 2.11 context:** A possible value of a specific instance of a *context variable*.
- 2.12 context variable:** A variable used in the *parsing process* to select which data structure is to be used for the *adaptive VLC* parsing of a given syntax element.
- 2.13 DC coefficient:** The first subset when the *transform coefficients*, that are contained in a specific *macroblock* and a specific *component*, are partitioned into 3 subsets.
- 2.14 DC-LP array:** The array of all *DC coefficients* and *low-pass coefficients*, for all *macroblocks* associated with a specific *component*.
- 2.15 decoder:** An embodiment of a *parsing process* and *decoding process*.
- 2.16 decoding process:** The process of computing output sample values from the parsed syntax elements of the *codestream*.
- 2.17 dequantization:** The process of rescaling the quantized *transform coefficients* after their value has been parsed from the *codestream* and before they are presented to the *inverse transform process*.
- 2.18 encoder:** An embodiment of an *encoding process*.
- 2.19 encoding process:** The process of converting source sample values into a *codestream*.
- 2.20 file:** A finite-length sequence of *bytes* that is accessible to a *decoder* in a manner such that the *decoder* can obtain access to the data at specified positions within the sequence of *bytes* (e.g., by storing the entire sequence of *bytes* in random access memory or by performing "position seek" operations to specified positions within the sequence of *bytes*).
- 2.21 file format:** A specified structure for the content of a *file*.
- 2.22 fixed-length code (FLC):** A code which assigns a finite set of allowable bit patterns to a specific set of values, where each bit pattern has the same length.
- 2.23 FLC refinement:** The *fixed-length coded* part of a *DC coefficient* or *low-pass coefficient* that is parsed using *adaptive fixed-length codes*.
- 2.24 flexbits:** The *fixed-length coded* part of the *high-pass coefficient* information which is parsed using *adaptive fixed-length codes*.
- 2.25 frequency band:** A collective term for one of the following three subsets of the *transform coefficients* for an *image*, which are separately parsed: *DC coefficients*, *low-pass coefficients*, and *high-pass coefficients*.
- 2.26 frequency mode:** A *codestream* structure mode where the DC, low-pass, high-pass and *flexbits frequency bands* for each *tile* are grouped separately.
- 2.27 hard tiles:** A *codestream* structure mode where the overlap operators are not applied across tile boundaries. Instead, boundary overlap operators are applied at tile boundaries.
- 2.28 high-pass coefficients:** The third subset, when the *transform coefficients* that are contained in a specific *macroblock* and a specific *component* are partitioned into 3 subsets.
- 2.29 image:** The result of the *decoding process*, consisting of a *primary image plane* and an optional *alpha image plane*.
- 2.30 image plane:** A collective term for a grouping of the *components* of the *image*.
- 2.31 internal colour format:** The colour format associated with the spatial-domain samples obtained through the *inverse transform process* and the *sample reconstruction process*, and distinguished from the *output colour format* associated with the *output formatting process*.
- 2.32 inverse core transform (ICT):** The two steps of the *inverse transform process* that involve processing of *transform coefficients* associated with each *macroblock* independently, with no *overlap filtering*.

- 2.33 inverse transform process:** The part of the *decoding process* by which a set of *dequantized transform coefficients* are converted into spatial-domain values.
- 2.34 inverse scanning:** The process of reordering an ordered set of parsed *syntax elements* from the *codestream* to form an array of *transform coefficients* associated with a specific *component* and *macroblock*.
- 2.35 low-pass coefficients:** The second subset, when the *transform coefficients* that are contained in a specific *macroblock* and a specific *component* are partitioned into 3 subsets.
- 2.36 luma:** The *component* of an *image plane* with index zero, and the *transform coefficients* and sample values associated with this *component*. Although this term is commonly associated with a signal that conveys perceptual brightness information, as used in this Specification the term is primarily an identifier of a particular array of samples or *transform coefficients* for an *image*.
- 2.37 macroblock:** The collection of *transform coefficients* or samples, across all *components*, that have the same indices *i* and *j* with respect to a *macroblock partition*.
- 2.38 macroblock partition:** The partitioning of each *component*, into 16×16, 8×8, or 16×8 *blocks*, depending on the *internal colour format*.
- 2.39 output bit depth:** The representation, including the number of bits and the interpretation of the bit pattern, used for the sample values of the output *image* that are the result of the *decoding process*.
- 2.40 output colour format:** The colour format associated with the output *image* that is the result of the *decoding process*.
- 2.41 output formatting process:** The process of converting the arrays of samples – that are the result of the *sample reconstruction process* – into the output samples that constitute the output of the *decoding process*. This specifies a conversion (if necessary) into the appropriate *output colour format* and *output bit depth*.
- 2.42 overlap filtering:** The steps of the *inverse transform process* that involve processing of *transform coefficients* across adjacent *blocks* and *macroblocks*.
- NOTE – When *overlap filtering* is applied, it is applied across *macroblock* boundaries as well as *block* boundaries. When the *codestream* uses *soft tiles*, the *overlap filtering* is also applied across *tile* boundaries. Otherwise, *overlap filtering* does not occur across *tile* boundaries.
- 2.43 parsing process:** The process of extracting bit sequences from the *codestream*, converting these bit sequences to syntax element values, and setting the values of global variables for use in the *decoding process*.
- 2.44 prediction:** The process of computing an estimate of the sample value or data element that is currently being decoded.
- 2.45 primary image plane:** The *image plane* that consists of all *image components* that are not a part of the *alpha image plane*.
- 2.46 quantization parameter (QP):** A value used to compute the scaling factor for the *dequantization* of a *transform coefficient*, before the *inverse transform process* is applied.
- 2.47 raster scan order:** The scan order in which a two-dimensional array of values is scanned row-wise from left to right, and the rows are scanned from the top row to the bottom.
- 2.48 refinement:** The process of modifying a predicted or partially-computed *transform coefficient*.
- 2.49 run:** The number of zero valued coefficient levels that precede a non-zero valued coefficient level in the *zigzag scan order* during the *inverse scanning process*.
- 2.50 sample reconstruction process:** The process of converting *dequantized transform coefficients* into samples of the *image*.
- 2.51 soft tiles:** A *codestream* structure mode where the overlap operators are applied across *tile* boundaries.
- 2.52 spatial mode:** A *codestream* structure mode where the *DC*, *low-pass*, *high-pass* and *flexbits frequency bands* for each specific *macroblock* are grouped together.
- 2.53 spatial transformation:** An element in the *codestream* indicating the preferred final displayed orientation of the decoded *image*, as specified in Rec. ITU-T T.832 | ISO/IEC 29199-2. The *spatial transformation* is only a suggestion, and *decoder conformance* is checked only for the decoded *image* prior to the application of this transformation (i.e., for orientation 0).
- 2.54 start code:** A bit pattern that specifies the beginning of a *tile packet* or other distinguished, contiguous set of syntax elements in the *codestream*.