FINAL DRAFT

INTERNATIONAL STANDARD

ISO/IEC FDIS 21794-2

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

Secretariat: JISC

Voting begins on:
2021-01-04

Voting terminates on:
2021-03-01

Information technology — Plenoptic image coding system (JPEG Pleno) —

Part 2: Light field coding

Reference number
ISO/IEC FDIS 21794-2:2021(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).

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) or the IEC list of patent declarations received (see patents.iec.ch).

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.

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 21794 series can be found on the ISO website.

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.

Introduction

This document is part of a series of standards for a system known as JPEG Pleno. This document defines the JPEG Pleno framework. It facilitates the capture, representation, exchange and visualization of plenoptic imaging modalities. A plenoptic image modality can be a light field, point cloud or hologram, which are sampled representations of the plenoptic function in the form of, respectively, a vector function that represents the radiance of a discretized set of light rays, a collection of points with position and attribute information, or a complex wavefront. The plenoptic function describes the radiance in time and in space obtained by positioning a pinhole camera at every viewpoint in 3D spatial coordinates, every viewing angle and every wavelength, resulting in a 7D function.

JPEG Pleno specifies tools for coding these modalities while providing advanced functionality at system level, such as support for data and metadata manipulation, editing, random access and interaction, protection of privacy and ownership rights.

Information technology — Plenoptic image coding system (JPEG Pleno) —

Part 2: Light field coding

1 Scope

This document specifies a coded codestream format for storage of light field modalities as well as associated metadata descriptors that are light field modality specific. This document also provides information on the encoding tools.

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.

ITU-T Rec. T.800 | ISO/IEC 15444-1, *Information technology — JPEG 2000 image coding system — Part 1: Core coding system*

ITU-T Rec. T.801 | ISO/IEC 15444-2, *Information technology — JPEG 2000 image coding system — Part 2: Extensions*

ISO/IEC 21794-1:2020, *Information technology — Plenoptic image coding system (JPEG Pleno) — Part 1: Framework*

ISO/IEC 60559, *Information technology — Microprocessor Systems — Floating-Point arithmetic*

3 Terms and definitions

For the purposes of this document the terms and definitions given in ISO/IEC 21794-1 and the following apply.

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

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

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

3.1
arithmetic coder
entropy coder that converts variable length strings to variable length codes (encoding) and vice versa (decoding)

3.2
bit-plane
two-dimensional array of bits

3.3
4D bit-plane
four-dimensional array of bits

3.4
coefficient
numerical value that is the result of a transformation or linear regression

3.5
compression
reduction in the number of bits used to represent source image data

3.6
depth
distance of a point in 3D space to the camera plane

3.7
disparity view
image that for each pixel of the subaperture view contains the apparent pixel shift between two subaperture views along either horizontal or vertical axis

3.8
hexadeca-tree
division of a 4D region into 16 (sixteen) 4D subregions

3.9
pixel
collection of sample values in the spatial image domain having all the same sample coordinates

EXAMPLE A pixel may consist of three samples describing its red, green and blue value.

3.10
plenoptic function
amount of radiance in time and in space by positioning a pinhole camera at every viewpoint in 3D spatial coordinates, every viewing angle and every wavelength, resulting in a 7D representation

3.11
reference view
subaperture view that is used as one of the references to generate the intermediate views

3.12
subaperture view
subaperture image
image taken of the 3D scene by a pinhole camera positioned at a particular viewpoint and viewing angle

3.13
texture
pixel attributes

EXAMPLE Colour information, opacity, etc.

3.14
transform
transformation
mathematical mapping from one signal space to another

4 Symbols and abbreviated terms

4.1 Symbols

Codestream_Body()	coded image data in the codestream without Codestream_Header()
Codestream_Header()	codestream header preceding the image data in the codestream
$\tilde{D}^{DEC}(t,s,v,u)$	decoded normalized disparity value at view (t,s) for pixel location (v,u)
$\tilde{D}(t,s,v,u)$	normalized disparity value at view (t,s) for pixel location (v,u)
$DPEC_k$	pointer to contiguous codestream for normalized disparity view k
D_{shift}	scaling parameter to translate quantized normalized disparity maps to positive range
DCODEC	disparity view codec type
f	focal length
FPW_p	fixed-weight merging parameter for view p
$H(t,s)$	view hierarchy value for view (t,s)
$HCC(t,s)$	horizontal camera centre coordinate for view (t,s)
$H_D(t,s)$	binary value defining the availability of a normalized disparity view (t,s)
J_0	Lagrangian encoding cost
J_1	Lagrangian encoding cost of spatial partitioning
J_2	Lagrangian encoding cost of view partitioning
$KR_{p,c}$	sparse filter regressor mask of texture component c for view p
LightField()	JPEG Pleno light field codestream
$LSW_j^{p,c}$	quantized least-squares merging weight of texture component c for view p, $j=1,2,\ldots,NLS_p$
MIDV	absolute value of the minimum value over all quantized normalized disparity views

$MMODE_p$	view merging mode for intermediate view p
MSP_p	sparse filter order for view p
NLS_p	number of least-squares merging coefficients for intermediate view p
NRT_p	regressor template size parameter for sparse filter for view p
NC	number of components in an image
N_I	number of intermediate views
N_{NDV}	number of reference normalized disparity views
N_p^D	number of normalized disparity reference views for intermediate view p
N_p^T	number of texture reference views for intermediate view p
N_{REF}	number of reference views
N_{RES}	number of prediction residual views
N_{sp}	total available number of regressors for sparse filter
Plev	level a particular codestream complies to
Ppih	profile a particular codestream complies to
Q_p	2D image of dimensions $V \times U$, defines the occlusion state-based segmentation at Intermediate view p
Q	normalized disparity quantization parameter
R	rate or bitrate, expressed in bit per sample
RCODEC	prediction residual view codec type
RDATA	array of bytes containing for a single prediction residual view the RCODEC codestream after header information has been stripped
RENCODING	array of bytes containing for a single prediction residual view the full DCODEC codestream
RGB	colour data for the red, green and blue colour component of a pixel
RHEADER	array of bytes containing for a single prediction residual view the header information from the RCODEC codestream

$RPEC_j$	pointer to contiguous codestream for prediction residual view j
s	coordinate of the addressed subaperture image along the s-axis
S	size of the light field image along the s-axis (COLUMNS)
s_{ii}^{Tr}	subscript of the column index of the reference view, $ii = 1, 2, \ldots, N_p^T$ in the light field array in row-wise scanning order
s_{jj}^{Dr}	subscript of the column index of the reference normalized disparity view, $jj = 1, 2, \ldots, N_p^D$ in the light field array in row-wise scanning order
SF_p	binary variable, determines if sparse filter is used (true) or not (false)
$SPW_j^{p,0}$	quantized sparse filter coefficients of texture component c for view p, $j = 1, 2, \ldots, MSP_p$
$\widehat{SPW}_j^{p,c}$	de-quantized sparse filter coefficients of texture component c for view p, $j = 1, 2, \ldots, MSP_p$
t	coordinate of the addressed subaperture image along the t-axis
T	size of the light field image along the t-axis (ROWS)
t_{ii}^{Tr}	subscript of the row index of the reference view, $ii = 1, 2, \ldots, N_p^T$ in the light field array in row-wise scanning order
t_{jj}^{Dr}	subscript of the row index of the reference normalized disparity view, $jj = 1, 2, \ldots, N_p^D$ in the light field array in row-wise scanning order
$\left(t_k^D, s_k^D\right)$	view coordinate subscripts for normalized disparity view k
$\left(t_l^X, s_l^X\right)$	view coordinate subscripts for reference view l
$\left(t_p^I, s_p^I\right)$	view coordinate subscripts for intermediate view p
$t_k \times s_k \times v_k \times u_k$	4D block dimensions at the 4D block partitioning stage
$t_b \times s_b \times v_b \times u_b$	4D block dimensions at the bit-plane hexadeca-tree decomposition stage
TCODEC	reference view codec type
TDATA	array of bytes, containing for a single reference view, the TCODEC codestream, after header information has been stripped

TENCODING	array of bytes, containing for a single reference view the full TCODEC codestream
THEADER	array of bytes, containing for a single reference view the header information from the TCODEC codestream
$TPEC_l$	pointer to contiguous codestream for reference view l
u	sample coordinate along the u-axis within the addressed subaperture image
U	size of the subaperture image along the u-axis (WIDTH)
v	sample coordinate along the v-axis within the addressed subaperture image
V	size of the subaperture image along the v-axis (HEIGHT)
$VCC\left(t,s\right)$	vertical camera centre coordinate for view $\left(t,s\right)$
VPP_p	view prediction parameters for intermediate view p
$X\left(t,s,v,u,c\right)$	texture value at view $\left(t,s\right)$ for pixel location $\left(v,u\right)$ for texture component c
$X^{DEC}\left(t,s,v,u,c\right)$	decoded texture value at view $\left(t,s\right)$ for pixel location $\left(v,u\right)$ for texture component c
$X_W^{\left(t_1,s_1\right)}\left(t_2,s_2\right)$	result of warping the texture view $\left(t_1,s_1\right)$ to view location $\left(t_2,s_2\right)$
Δx	horizontal distance between a pair of camera centres
Δy	vertical distance between a pair of camera centres
YCbCr	colour data for the luminance, the blue chrominance and the red chrominance component of a pixel
$z\left(t,s,v,u\right)$	depth value at view $\left(t,s\right)$ for pixel location $\left(v,u\right)$
$\hat{\theta}_i^p$	distance based merging weight for reference view $i=1,\ldots,N_p^T$ at intermediate view p
α_i^p	distance based factor, used for defining the merging weight, at intermediate view p for reference view $i=1,\ldots,N_p^T$
Γ_p	binary matrix, defining the locations of the non-zero merging weights in merging weight matrix $\Theta_{p,c}$ at intermediate view p. It is identical between all colour components c

$\theta_j^{p,c}$ de-quantized least-squares merging weight of texture component c for view p, $j = 1, 2, \ldots, NLS_p$

$\theta_{p,c}^{sp}$ sparse filter coefficients at intermediate view p for colour component c

$\Theta_{p,c}$ merging weight matrix for intermediate view p for colour component c

$Y_{p,c}$ locations of the non-zero elements of $\Psi_{(v,u)}$

$\Psi_{(v,u)}$ regressor template at pixel location (v, u)

Ω_p^{Dr} set of reference normalized disparity views for intermediate view p

Ω_p^{occlD} set of occluded pixels, which remain to be inpainted, during normalized disparity view synthesis at intermediate view p

Ω_p^{occlT} set of occluded pixels, which remain to be inpainted, during texture view synthesis at intermediate view p

Ω_p^{Tr} set of reference views for intermediate view p

4.2 Abbreviated terms

2D two dimensional

3D three dimensional

4D four dimensional

DCT discrete cosine transform

floating point floating point notation as specified in ISO/IEC 60559

HTTP hypertext transfer protocol

IDCT inverse DCT

IPR intellectual property rights

IV intermediate view; subaperture view that is generated from surrounding reference view(s)

JPEG Joint Photographic Experts Group

JPL JPEG Pleno file format

LSB least significant bit

MSB most significant bit

R-D	rate-distortion
RV	reference view
URL	uniform resource locator
XML	eXtensible Markup Language

5 Conventions

5.1 Naming conventions for numerical values

Integer numbers are expressed as bit patterns, hexadecimal values or decimal numbers. Bit patterns and hexadecimal values have both a numerical value and an associated particular length in bits.

Hexadecimal notation, indicated by prefixing the hexadecimal number by "0x", may be used instead of binary notation to denote a bit pattern having a length that is an integer multiple of 4. For example, 0x41 represents an eight-bit pattern having only its second most significant bit and its least significant bit equal to 1. Numerical values that are specified under a "**Code**" heading in tables that are referred to as "code tables" are bit pattern values (specified as a string of digits equal to 0 or 1 in which the left-most bit is considered the most-significant bit). Other numerical values not prefixed by "0x" are decimal values. When used in expressions, a hexadecimal value is interpreted as having a value equal to the value of the corresponding bit pattern evaluated as a binary representation of an unsigned integer (i.e. as the value of the number formed by prefixing the bit pattern with a sign bit equal to 0 and interpreting the result as a two's complement representation of an integer value). For example, the hexadecimal value 0xF is equivalent to the 4-bit pattern '1111' and is interpreted in expressions as being equal to the decimal number 15.

5.2 Operators

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

5.2.1 Arithmetic operators

+	addition
−	subtraction (as a binary operator) or negation (as a unary prefix operator)
×	multiplication
/	division without truncation or rounding
<<	left shift; x<<s is defined as $x \times 2^s$
>>	right shift; x>>s is defined as $\lfloor x/2^s \rfloor$
++	increment with 1
--	decrement with 1

umod	x umod a is the unique value y between 0 and a–1 for which y+Na = x with a suitable integer N
&	bitwise AND operator; compares each bit of the first operand to the corresponding bit of the second operand If both bits are 1, the corresponding result bit is set to 1. Otherwise, the corresponding result bit is set to 0.
^	bitwise XOR operator; compares each bit of the first operand to the corresponding bit of the second operand If both bits are equal, the corresponding result bit is set to 0. Otherwise, the corresponding result bit is set to 1.

5.2.2 Logical operators

||	logical OR
&&	logical AND
!	logical NOT

5.2.3 Relational operators

>	greater than
>=	greater than or equal to
<	less than
<=	less than or equal to
==	equal to
!=	not equal to

5.2.4 Precedence order of operators

Operators are listed 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
[]	indexing of arrays	left to right
++, --	increment, decrement	left to right
!, –	logical not, unary negation	