INTERNATIONAL STANDARD

ISO/IEC 15938-13

First edition
2015-09-01

Information technology — Multimedia content description interface —

Part 13: Compact descriptors for visual search

Technologies de l'information — Interface de description du contenu multimédia —

Partie 13: Descripteurs compacts pour recherche visuelle

Reference number
ISO/IEC 15938-13:2015(E)

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 15938 consists of the following parts, under the general title *Information technology — Multimedia content description interface*:

— *Part 1: Systems*

— *Part 2: Description definition language*

— *Part 3: Visual*

— *Part 4: Audio*

— *Part 5: Multimedia description schemes*

— *Part 6: Reference software*

— *Part 7: Conformance testing*

— *Part 8: Extraction and use of MPEG-7 descriptions*

— *Part 9: Profiles and levels*

— *Part 10: Schema definition*

— *Part 11: MPEG-7 profile schemas*

— *Part 12: Query format*

— *Part 13: Compact descriptors for visual search*

Introduction

This International Standard, also known as "Multimedia Content Description Interface," provides a standardized set of technologies for describing multimedia content. It addresses a broad spectrum of multimedia applications and requirements by providing a metadata system for describing the features of multimedia content.

The following are specified in this International Standard:

— **Description schemes (DS)** describe entities or relationships pertaining to multimedia content. Description schemes specify the structure and semantics of their components, which may be Description Schemes, descriptors, or datatypes.

— **Descriptors (D)** describe features, attributes, or groups of attributes of multimedia content.

— **Datatypes** are the basic reusable datatypes employed by description schemes and descriptors.

— **Systems tools** support delivery of descriptions, multiplexing of descriptions with multimedia content, synchronization, file format, and so forth.

This International Standard is subdivided into 13 parts:

— **Part 1 — Systems**: specifies the tools for preparing descriptions for efficient transport and storage, compressing descriptions, and allowing synchronization between content and descriptions.

— **Part 2 — Description definition language**: specifies the language for defining the International Standard set of description tools (DSs, Ds, and datatypes) and for defining new description tools.

— **Part 3 — Visual**: specifies the description tools pertaining to visual content.

— **Part 4 — Audio**: specifies the description tools pertaining to audio content.

— **Part 5 — Multimedia description schemes**: specifies the generic description tools pertaining to multimedia including audio and visual content.

— **Part 6 — Reference software**: provides a software implementation of the International Standard.

— **Part 7 — Conformance testing**: specifies the guidelines and procedures for testing conformance of implementations of the International Standard.

— **Part 8 — Extraction and use of MPEG-7 descriptions**: provides guidelines and examples of the extraction and use of descriptions.

— **Part 9 — Profiles and levels**: provides guidelines and standard profiles.

— **Part 10 — Schema definition**: specifies the schema using description definition language.

— **Part 11 — Profile Schemas**: listing of profile schemas using description definition language.

— **Part 12 — Query format**: contains the tools of the MPEG Query Format (MPQF).

— **Part 13 — Compact descriptors for visual search**: specifies an image description tool for visual search applications.

Information technology — Multimedia content description interface —

Part 13: Compact descriptors for visual search

1 Scope

The structure of this part of ISO/IEC 15938 is as follows. Clauses 2 and 3 specify the terms, abbreviations, symbols, and conventions used in the International Standard. Clause 4 specifies the binary representation syntax and descriptor component semantics for a CDVS image descriptor. Clause 5 specifies the extraction and encoding process for a CDVS image descriptor. Annexes A-J specify information relevant to the encoding process of Clause 5. Annex K contains an informative description of the decoding process of a CDVS image descriptor.

This part of the MPEG-7 standard specifies an image description tool designed to enable efficient and interoperable visual search applications, allowing visual content matching in images. Visual content matching includes matching of views of objects, landmarks, and printed documents, while being robust to partial occlusions as well as changes in viewpoint, camera parameters, and lighting conditions.

2 Terms and definitions

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

2.1
image descriptor
descriptor extracted from one image

2.2
image descriptor length
size of an image descriptor in bytes

Note 1 to entry: This International Standard specifies six average (i.e. over a large number of images) image descriptor lengths, i.e. 512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, and 16384 bytes, and the encoding process for each image descriptor length.

2.3
original image
input image to the image descriptor encoder

2.4
converted image
image which is a spatially resampled version of the original image and from which the image descriptor is extracted

2.5
pixel
indexable element of the original image or the converted image, comprising spatial coordinates and a luminance value

2.6
interest point
point in an image showing detection stability under local and global perturbations in the image domain, including perspective transformations, changes in image scale, and illumination variations

2.7
local region
area in an image in the neighbourhood of an interest point, used to generate local feature descriptors

2.8
cell
each of the 4x4 subdivisions of a local region

2.9
cell histogram
histogram of gradients computed from the cell

2.10
local feature descriptor
descriptor of a local region, computed from the cell histograms

2.11
global descriptor
aggregation of local feature descriptors into a compact representation of the image

2.12
compressed local feature descriptor
compressed representation of a local feature descriptor

2.13
interest point coordinate
horizontal and vertical pixel coordinates indicating the position of an interest point in the converted image resolution, rounded to the nearest integer

2.14
location quantization factor
size of the blocks of the spatial grid superimposed on top of the converted image in order to obtain quantized interest point coordinates' values

2.15
histogram map
binary representation of the converted image scaled down by the location quantization factor, indicating whether each bin generated through the superimposition of the spatial grid on top of the converted image is populated with at least one interest point

2.16
histogram count
vector indicating the number of interest points that populate each non-empty bin generated through the superimposition of a spatial grid on top of the converted image

3 Symbols and abbreviated terms

3.1 General

NOTE The mathematical operators used in this part of ISO/IEC 15938 are similar to those used in the C programming language. Unless otherwise indicated, all the arithmetic operations are performed with real values. Numbering and counting conventions generally begin from 0.

3.2 Abbreviations

CDVS Compact Descriptors for Visual Search

LoG Laplacian-of-Gaussian

MPEG Moving Picture Experts Group

MPEG-7 ISO/IEC 15938

3.3 Arithmetic operations

+ Addition

- Subtraction (as a binary operator) or negation (as a unary operator)

++ Increment by 1, i.e. x++ is equivalent to x=x+1

-- Decrement by 1, i.e. x-- is equivalent to x=x-1

+= Increment by value, i.e. x+=y is equivalent to x=x+y

-= Decrement by value, i.e. x-=y is equivalent to x=x-y

***** Multiplication (in binary representation syntax and pseudo-code) or convolution (elsewhere)

× Multiplication

· Multiplication

/ Division

÷ Division

% Modulo operator

3.4 Logical operators

|| Logical OR

v Logical OR

&& Logical AND

∧ Logical AND

! Logical NOT

3.5 Relational operators

> Greater than

>= Greater than or equal to

≥ Greater than or equal to

< Less than

<= Less than or equal to

≤ Less than or equal to

== Equal to

!= Not equal to

3.6 Bitwise operators

|	OR
&	AND

3.7 Assignment

=	Assignment operator
←	Assignment operator

3.8 Mnemonics

The following mnemonics are defined to describe the different data types used in the coded bitstream.

bslbf	Bit string, left bit first, where "left" is the order in which bits are written in the bitstream.
uimsbf	Unsigned integer, most significant bit first.
vlclbf	Variable length code, left bit first, where "left" refers to the order in which the VLC codes are written in the bitstream and where the byte order of multibyte words is most significant byte first.

3.9 Constants

π	3.141 592 653 58...
e	2.718 281 828 45...

3.10 Functions

$\log_n(\,)$	Base-n logarithm
max( )	Maximum value in argument list
min( )	Minimum value in argument list
sgn( )	Sign function, i.e. sgn(x) = -1, 0 or +1 when x < 0, x == 0 or x > 0, respectively
$\lvert\ \rvert$	Absolute value of scalar or a vector norm
$\lfloor\ \rfloor$	Floor function which returns the maximum integer number less than or equal to the given real number
$\lceil\ \rceil$	Ceiling function which returns the minimum integer number greater than or equal to the given real number
$\downarrow_{2x2}$	Downsamples an image by keeping only the even rows and even columns of the image, without anti-alias filtering

4 CDVS syntax

4.1 Binary representation syntax

CDVSDescriptor {	Number of bits	Mnemonics
VersionID	3	bslbf
ModeID	8	uimsbf
GlobalHasBitSelection	1	bslbf
GlobalHasVariance	1	bslbf
RelevanceBitsPresent	1	bslbf
ReservedBits	2	bslbf
OriginalImageXResolution	16	uimsbf
OriginalImageYResolution	16	uimsbf
NumberOfLocalDescriptors	16	uimsbf
if(NumberOfLocalDescriptors>0) {		
for(k=0; k<NumberOfGlobalFunctions; k++) {		
GlobalFunctionPresent[k]	1	bslbf
}		
if(GlobalHasBitSelection) {		
for(k=0; k<NumberOfGlobalFunctions; k++) {		
if(GlobalFunctionPresent[k]) {		
GlobalFunctionMeanVector[k]	24	bslbf
}		
}		
}		
else {		
for(k=0; k<NumberOfGlobalFunctions; k++) {		
if(GlobalFunctionPresent[k]) {		
GlobalFunctionMeanVector[k]	32	bslbf
}		
}		
}		
if(GlobalHasVariance) {		
for(k=0; k<NumberOfGlobalFunctions; k++) {		
if(GlobalFunctionPresent[k]) {		
GlobalFunctionVarianceVector[k]	32	bslbf
}		
}		
}		
HistogramCountSize	16	uimsbf
HistogramMapSizeX	16	uimsbf
HistogramMapSizeY	16	uimsbf
HistogramCount (arithmetically coded block; see 5.8)	>=0	vlclbf

CDVSDescriptor {	Number of bits	Mnemonics
HistogramMap (arithmetically coded block; see 5.8)	>=0	vlclbf
NumberOfElementGroups	6	uimsbf
for(k=0; k<NumberOfLocalDescriptors; k++) {		
for(n=0; n<(4*NumberOfElementGroups); n++) {		
LocalDescriptorElements[k][n]	1-2	vlclbf
}		
}		
if(RelevanceBitsPresent) {		
for(k=0; k<NumberOfLocalDescriptors; k++)		
RelevanceBits[k]	1	bslbf
}		
}		
BitStuffing	0-7	vlclbf
}		
}		

VersionID = 1

NumberOfGlobalFunctions = 512

4.2 Descriptor component semantics

VersionID

This descriptor component specifies the CDVSDescriptor version. In this International Standard ISO/IEC 15938-13:2015, VersionID = 1.

ModeID

This descriptor component specifies the image descriptor length. There are six image descriptor lengths, and their corresponding ModeID values are shown in Table 1 below.

Table 1 — ModeID values for the six image descriptor lengths

Image descriptor length	ModeID
512 bytes	1
1024 bytes	2
2048 bytes	3
4096 bytes	4
8192 bytes	5
16384 bytes	6

GlobalHasBitSelection

This descriptor component specifies whether bit selection is applied or not to the GlobalFunctionMeanVector of each of the Gaussian functions which are present in the global descriptor of an image descriptor. If GlobalHasBitSelection == 1 then bit selection is applied, and if GlobalHasBitSelection == 0 then bit selection is not applied. More details are provided in 5.6.

GlobalHasVariance

This descriptor component specifies whether the GlobalFunctionVarianceVector of each of the Gaussian functions which are present in the global descriptor of an image descriptor appears in the bitstream or not. If GlobalHasVariance == 1 then GlobalFunctionVarianceVector appears in the bitstream, and if GlobalHasVariance == 0 then GlobalFunctionVarianceVector does not appear in the bitstream. More details are provided in 5.6.

RelevanceBitsPresent

This descriptor component specifies if a relevance bit for each compressed local feature descriptor is present in the bitstream. If RelevanceBitsPresent == 1 then the relevance bits are present in the bitstream, and if RelevanceBitsPresent == 0 then the relevance bits are not present in the bitstream. More details are provided in 5.4.

ReservedBits

This descriptor component comprises two bits which are reserved for future use and they shall both be set to 0.

OriginalImageXResolution

This descriptor component specifies the width (in pixels) of the original image.

OriginalImageYResolution

This descriptor component specifies the height (in pixels) of the original image.

NumberOfLocalDescriptors

This descriptor component specifies the number of compressed local feature descriptors which are present in the bitstream. More details are provided in 5.10. NumberOfLocalDescriptors == 0 indicates that no local features were identified in the image.

NumberOfGlobalFunctions

This descriptor component specifies the maximum number of Gaussian functions used in the global descriptor and has a value NumberOfGlobalFunctions = 512. More details are provided in 5.6.

GlobalFunctionPresent

This descriptor component specifies a 1-D array of size NumberOfGlobalFunctions indicating which Gaussian functions are present in the global descriptor of a particular image descriptor. If a Gaussian function is present in the global descriptor the corresponding value in the array is 1, otherwise it is 0. More details are provided in 5.6.

GlobalFunctionMeanVector

This descriptor component specifies a 1-D array of size equal to the number of Gaussian functions which are present in the global descriptor, i.e. those Gaussian functions with a corresponding value of 1 in GlobalFunctionPresent. Each entry in the array is the binarized mean vector of the corresponding global descriptor Gaussian function, and the length of each vector is 24 bits if GlobalHasBitSelection == 1 and 32 bits if GlobalHasBitSelection == 0. More details are provided in 5.6.

GlobalFunctionVarianceVector

This descriptor component specifies a 1-D array of size equal to the number of Gaussian functions which are present in the global descriptor, i.e. those Gaussian functions with a corresponding value of 1 in GlobalFunctionPresent. Each entry in the array is the binarized variance vector of the corresponding global descriptor Gaussian function. More details are provided in 5.6.

HistogramCountSize

This descriptor component specifies the histogram count vector length for location coding. More details are provided in 5.8.

HistogramMapSizeX

This descriptor component specifies the horizontal x resolution of the histogram map for location coding. More details are provided in 5.8.

HistogramMapSizeY

This descriptor component specifies the vertical y resolution of the histogram map for location coding. More details are provided in 5.8.

HistogramCount

This descriptor component specifies a vector for location coding, containing the number of non-zero elements for each non-null block of the histogram map. More details are provided in 5.8.

HistogramMap

This descriptor component specifies a 2D-array for location coding, containing a block representation of the converted image. Each block can assume a binary value, indicating the occurrence or not of interest points within that block. The array is scanned according a procedure described in 5.8. The scanning terminates when all the non-null elements of the Histogram Map are encoded. More details are provided in 5.8.

NumberOfElementGroups

This descriptor component specifies the number of element groups in each compressed local feature descriptor. Each element group contains four elements and the number of elements in each compressed local feature descriptor is given by 4×NumberOfElementGroups. More details are provided in 5.7.

LocalDescriptorElements

This descriptor component specifies a 2-D array of compressed local feature descriptor elements. The size of the first dimension is NumberOfLocalDescriptors and the size of the second dimension is 4×NumberOfElementGroups. LocalDescriptorElements[k][n] is the n^{th} element of the k^{th} compressed local feature descriptor. For each compressed local feature descriptor, its elements are ordered as described in 5.7.

The compressed local feature descriptors themselves are ordered as described in 5.9.

RelevanceBits

This descriptor component specifies a 1-D array of size NumberOfLocalDescriptors indicating which compressed local feature descriptors correspond to the top 300 local features as determined in 5.4. If the k^{th} local feature is one of the top 300 local features, then RelevaceBits[k] is set to 1, otherwise it is set to 0. If NumberOfLocalDescriptor<300, then all the values in RelevanceBits are set to 1. More details are provided in 5.4.

The relevance bits are ordered in the same order as the descriptors in LocalDescriptorElement, as described in 5.9.

BitStuffing

This descriptor component specifies stuffing bits (a sequence of '1's) to align the descriptor to a byte boundary.

5 CDVS encoding

5.1 General

This clause specifies the encoder operations for computing an image descriptor. A simplified diagram of a complete CDVS encoder implementing these encoding operations is presented in informative Annex A.

5.2 Original image preprocessing

The original image is a luminance raster image containing values in the interval [0, 255] where increasing values correspond to increasing luminance. The exact mapping of luminance values within this interval is beyond the scope of the standard. If at least one of the dimensions of the original image is greater than 640 pixels then the original image shall be spatially resampled, maintaining the aspect ratio, so that the largest of the vertical and horizontal image dimensions is equal to 640 pixels, to obtain a converted image $J(x, y)$, in which $x \in \{0,\ldots,X-1\}$ and $y \in \{0,\ldots,Y-1\}$ are the horizontal and vertical pixel coordinates respectively, X and Y the pixel horizontal and vertical image dimensions respectively, and with coordinates (0,0) located at the top left corner of the image. For this resampling operation, a Lanczos filter with a = 3 should be used. If both the dimensions of the original image are no greater than 640 pixels, then no spatial resampling is performed and the content of the converted image shall be the same as the content of the original image.

5.3 Interest point detection

5.3.1 Introduction

This operation is performed using the ALP (A Low-degree Polynomial) detector. In order to find interest points, ALP approximates the result of the LoG filtering by means of polynomials, used to find extrema in the scale space and to refine the spatial position of the detected points.

5.3.2 Scale space construction

Let g denote the Gaussian kernel in two dimensions with positive scale parameter σ

$$g(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

The filtering operations shall be done at 4 scales with values for the σ parameter in an exponentially increasing sequence

$$\sigma_k=\sigma_0\cdot 2^{\frac{k}{2}}, k=0,\ldots,3 \tag{2}$$

as provided in Table 2 below.

Table 2 — Values of the scale parameter

k	σ_k
0	1,600000
1	2,262742
2	3,200000
3	4,525483