INTERNATIONAL STANDARD

ISO/IEC 15938-3

First edition
2002-02-15

AMENDMENT 2
2006-04-01

Information technology — Multimedia content description interface —

Part 3: **Visual**

AMENDMENT 2: Perceptual 3D Shape Descriptor

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

Partie 3: Visuel

AMENDEMENT 2: Descripteur sensoriel de forme 3D

Reference number
ISO/IEC 15938-3:2002/Amd.2:2006(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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.

Amendment 2 to ISO/IEC 15938-3:2002 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This document preserves the sectioning of ISO/IEC 15938-3. The text and figures given in this Amendment are currently being considered as additions and/or modifications to those corresponding sections in ISO/IEC 15938-3.

Information technology — Multimedia content description interface —

Part 3: Visual

AMENDMENT 2: Perceptual 3D Shape Descriptor

Replace subclause 1.2 with:

1.2 Overview of Visual Description Tools

This part of ISO/IEC 15938 specifies tools for description of visual content, including still images, video and 3D models. These tools are defined by their syntax in DDL and binary representations and semantics associated with the syntactic elements. They enable description of the visual features of the visual material, such as color, texture, shape and motion, as well as localization of the described objects in the image or video sequence. An overview of the visual description tools is shown in Figure 1.

The basic structure description tools include five supporting tools of visual descriptions defined in clauses 6-11. They are categorized into two groups, descriptor containers and basic supporting tools. The former consists of three datatypes, GridLayout providing efficient representations of visual features on grids, TimeSeries representing temporal arrays of several descriptions, GofGopFeature describes representative descriptions over video segment, and MultipleView describing a 3D object using several pictures captured from different view angles. The latter contains two tools, Spatial2DcoordinateSystem used to specify the 2D coordinate system and TemporalInterpolation indicating the interpolation method between two samples on a time axis.

The remaining description tools, except for the FaceRecognition descriptor, are associated with visual features and are grouped into five feature categories: Color, Texture, Shape, Motion and Localization.

The color description tools include five color descriptors to represent different aspects of color features: representative colors (DominantColor), color distribution (ScalableColor), spatial distribution of colors (ColorLayout and ColorStructure) and perceptual feeling of illumination color (ColorTemperature). It also contains three supporting tools, ColorSpace and ColorQuantization used in DominantColor and IlluminationInvariantColor to extend four color descriptors, DominantColor, ScalableColor, ColorLayout and ColorStructure, to support illumination invariant similarity matching. An extension of ScalableColor to a group of frames or pictures (GoFGoPColor) is also included in this group. All the color descriptors can be extracted from arbitrarily shaped regions.

The texture description tools facilitate browsing (TextureBrowsing) and similarity retrieval (HomogeneousTexture and EdgeHistogram) using the texture of a still or moving image region. All the texture descriptors can be extracted from arbitrarily shaped regions.

The shape description tools include two descriptors that characterize different shape features of a 2D object or region. The RegionShape descriptor captures the distribution of all pixels within a region and the Contour Shape descriptor characterizes the shape properties of the contour of an object. The extension of RegionShape is also defined as ShapeVariation to describe temporal variation of shape over video segment. The Shape3D and Perceptual 3D Shape descriptors provide 3-dimensional shape information; the former represents an intrinsic shape characterization of 3D mesh models, and the latter represents part-based representation of a 3D object.

The motion description tools include four descriptors that characterize various aspects of motion. The CameraMotion descriptor specifies a set of basic camera operations such as, for example, panning and tilting.

The motion of a key point (pixel) from a moving object or region can be characterized by the MotionTrajectory descriptor. The ParametricMotion descriptor characterizes an evolution of an arbitrarily shaped region over time in terms of a 2D geometric transformation. Finally, the MotionActivity descriptor captures the pace of the motion in the sequence, as perceived by the viewer. All motion descriptors except for CameraMotion can be extracted from arbitrarily shaped regions.

The localization description tools can be used to indicate regions of interest in the spatial (RegionLocator) and spatio-temporal (SpatioTemporalLocator) domains.

The FaceRecognition descriptor and the Advance Face Recognition descriptor are not associated with any particular visual feature and can be used to describe a human face for applications requiring the matching and retrieval of face images.

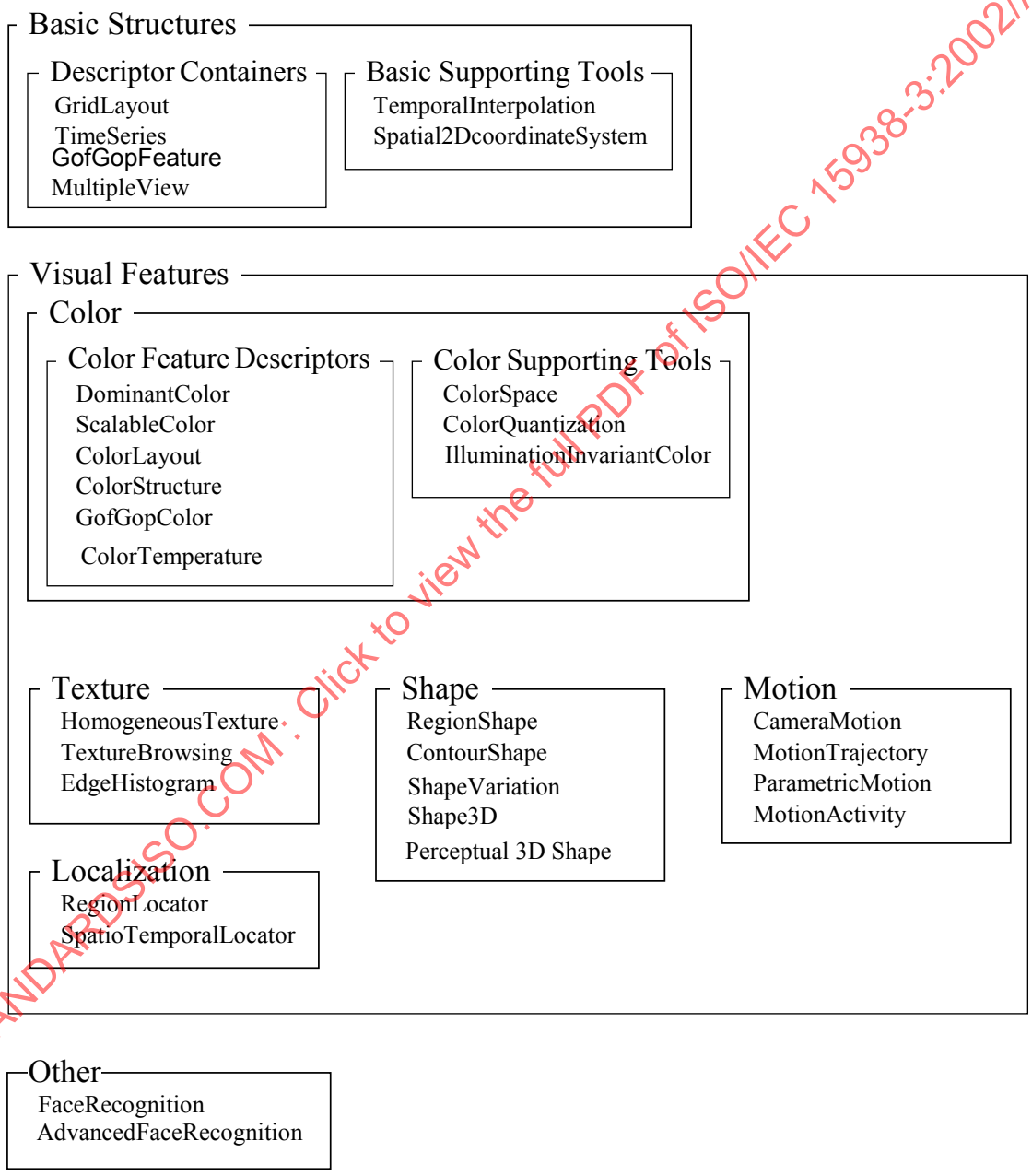

Figure 1 — Overview of Visual Description Tools

Extend the definitions in clause 2:

2.4 adjacency matrix

matrix with rows and columns labeled by graph vertices, with a 1 or 0 in position (i, j) according to whether *i*-th and *j*-th nodes are adjacent or not.

Replace subclause 4.2.2 with:

4.2.2 Generic binary representation

The use of the video-specific syntax is signalled using the codec configuration mechanism defined in ISO/IEC 15938-1. And the following classification scheme is defined for this purpose.

```
<ClassificationScheme uri="urn:mpeg:mpeg7:cs:VisualDescriptorCodecCS:2001">
   <Term termID="1">
      <Name xml:lang="en">MPEG7CameraMotion</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Camera Motion
            Codec</Definition>
   </Term>
   <Term termID="2">
      <Name xml:lang="en">MPEG7ColorLayout</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Color Layout
            Codec</Definition>
   </Term>
   <Term termID="3">
      <Name xml:lang="en">MPEG7ColorQuantization</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Color Quantization
            Codec</Definition>
   </Term>
   <Term termID="4">
      <Name xml:lang="en">MPEG7ColorSpace</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Color Space
            Codec</Definition>
   </Term>
   <Term termID="5">
      <Name xml:lang="en">MPEG7ColorStructure</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Color Structure
            Codec</Definition>
   </Term>
   <Term termID="6">
      <Name xml:lang="en">MPEG7ContourShape</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Contour Shape
            Codec</Definition>
   </Term>
   <Term termID="7">
      <Name xml:lang="en">MPEG7DominantColor</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Dominant Color
            Codec</Definition>
   </Term>
   <Term termID="8">
      <Name xml:lang="en">MPEG7EdgeHistogram</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Edge Histogram
            Codec</Definition>
   </Term>
   <Term termID="9">
      <Name xml:lang="en">MPEG7FaceRecognition</Name>
```

```
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Face Recognition
            Codec</Definition>
   </Term>
   <Term termID="10">
      <Name xml:lang="en">MPEG7FoFGoPColor</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary GoFGoP Color
            Codec</Definition>
   </Term>
   <Term termID="11">
      <Name xml:lang="en">MPEG7GridLayout</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Grid Layout
            Codec</Definition>
   </Term>
   <Term termID="12">
      <Name xml:lang="en">MPEG7HomogeneousTexture</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Homogeneous Texture
            Codec</Definition>
   </Term>
   <Term termID="13">
      <Name xml:lang="en">MPEG7IrregularVisualTimeSeries</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Irregular Time Series
            Codec</Definition>
   </Term>
   <Term termID="14">
      <Name xml:lang="en">MPEG7MotionActivity</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Motion Activity
            Codec</Definition>
   </Term>
   <Term termID="15">
      <Name xml:lang="en">MPEG7MotionTrajectory</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Motion Trajectory
            Codec</Definition>
   </Term>
   <Term termID="16">
      <Name xml:lang="en">MPEG7MultipleView</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Multiple View
            Codec</Definition>
   </Term>
   <Term termID="17">
      <Name xml:lang="en">MPEG7ParametricMotion</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Parametric Motion
            Codec</Definition>
   </Term>
   <Term termID="18">
      <Name xml:lang="en">MPEG7RegionLocator</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Region Locator
            Codec</Definition>
   </Term>
   <Term termID="19">
      <Name xml:lang="en">MPEG7RegionShape</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Region Shape
            Codec</Definition>
   </Term>
   <Term termID="20">
      <Name xml:lang="en">MPEG7RegularVisualTimeSeries</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Regular Time Series
            Codec</Definition>
   </Term>
   <Term termID="21">
      <Name xml:lang="en">MPEG7ScalableColor</Name>
```

```
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Scalable Color
            Codec</Definition>
   </Term>
   <Term termID="22">
      <Name xml:lang="en">MPEG7Shape3D</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Shape 3D
            Codec</Definition>
   </Term>
   <Term termID="23">
      <Name xml:lang="en">MPEG7Spatial2DCoordinateSystem</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Spatial 2D Coordinate
            System Codec</Definition>
   </Term>
   <Term termID="24">
      <Name xml:lang="en">MPEG7SpatioTemporalLocator</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary SpatioTemporal Locator
            Codec</Definition>
   </Term>
   <Term termID="25">
      <Name xml:lang="en">MPEG7TemporalInterpolation</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Temporal Interpolation
            Codec</Definition>
   </Term>
   <Term termID="26">
      <Name xml:lang="en">MPEG7TextureBrowsing</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Texture Browsing
            Codec</Definition>
   </Term>
   <Term termID="27">
      <Name xml:lang="en">MPEG7GofGopFeature</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Gof Gop Feature
Codec</Definition>
   </Term>
   <Term termID="28">
      <Name xml:lang="en">MPEG7ColorTemperature</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Color Temperature
Codec</Definition>
   </Term>
   <Term termID="29">
      <Name xml:lang="en">MPEG7ShapeVariation</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Shape Variation
Codec</Definition>
   </Term>
   <Term termID="30">
      <Name xml:lang="en">MPEG7IlluminationInvariantColor</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Illumination Invariant
Color Codec</Definition>
   </Term>
   <Term termID="31">
      <Name xml:lang="en">MPEG7AdvancedFaceRecognition</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Advanced Face Recognition
Codec</Definition>
   </Term>
   <Term termID="32">
      <Name xml:lang="en">MPEG7Perceptual3DShape</Name>
      <Definition xml:lang="en">ISO/IEC 15938-3 Binary Perceptual 3D Shape
Codec</Definition>
   </Term>
</ClassificationScheme>
```

Replace Table 1 in subclause 5.2.4 with:

Table 1 — Assignment of IDs to descriptors

ID	Descriptor
0	Forbidden
1	CameraMotion
2	ColorLayout
3	ColorSpace
4	ColorStructure
5	ColorQuantization
6	ContourShape
7	DominantColor
8	EdgeHistogram
9	FaceRecognition
10	GoFGoPColor
11	GridLayout
12	HomogeneousTexture
13	IrregularVisualTimeSeries
14	MotionActivity
15	MotionTrajectory
16	MultipleView
17	ParametricMotion
18	RegionLocator
19	RegionShape
20	RegularVisualTimeSeries
21	ScalableColor
22	Shape3D
23	Spatial2DCoordinateSystem
24	SpatioTemporalLocator
25	TemporalInterpolation
26	TextureBrowsing
27	GofGopFeature
28	ColorTemperature
29	ShapeVariation
30	IlluminationInvariantColor
31	AdvancedFaceRecognition
32	Perceptual3DShape
33-255	Reserved

Add after subclause 8.5:

8.6 Perceptual 3D Shape

The Perceptual 3D Shape descriptor is a part-based representation of a 3D object expressed as a graph. In this context “node” is a vertex in the graph representation corresponding to a part in the 3D model. Such a representation facilitates object description consistent with human perception. Compared with the Shape 3D descriptor, the Perceptual 3D Shape descriptor supports additional functionalities, like ‘Query by sketch’ and ‘Query by editing’, which make the content-based retrieval system more interactive and efficient in querying and retrieving similar 3D objects.

8.6.1 DDL representation syntax

```
<!-- ############################################################  -->
<!-- Definition of MPEG-7  P3DS Datatype                          -->
<!-- ############################################################  -->
<complexType name="Perceptual3DShapeType">
   <complexContent>
   <extension base=mpeg7:VisualDType>
   <sequence>
      <element name="BitsPerAttribute">
         <simpleType>
               <restriction base="integer">
               <minInclusive value="1"/>
               <maxInclusive value="15"/>
               </restriction>
         </simpleType>
      </element>
      <element name="IsConnected" minOccurs="0" maxOccurs="1">
         <simpleType>
            <restriction>
               <simpleType>
                  <list itemType="boolean"/>
               </simpleType>
               <minlength value="1"/>
               <maxlength value="32385"/>
            </restriction>
         </simpleType>
      </element>
      <sequence minOccurs="1" maxOccurs="255">
         <element name="Volume" type="mpeg7:unsigned15"/>
         <element name="Center_X" type="mpeg7:unsigned15"/>
         <element name="Center_Y" type="mpeg7:unsigned15"/>
         <element name="Center_Z" type="mpeg7:unsigned15"/>
         <element name="PCA_Axis_1X" type="mpeg7:unsigned15"/>
         <element name="PCA_Axis_1Y" type="mpeg7:unsigned15"/>
         <element name="PCA_Axis_1Z" type="mpeg7:unsigned15"/>
         <element name="PCA_Axis_2X" type="mpeg7:unsigned15"/>
         <element name="PCA_Axis_2Y" type="mpeg7:unsigned15"/>
         <element name="PCA_Axis_2Z" type="mpeg7:unsigned15"/>
         <element name="Max_1" type="mpeg7:unsigned15"/>
         <element name="Max_2" type="mpeg7:unsigned15"/>
         <element name="Max_3" type="mpeg7:unsigned15"/>
         <element name="Convexity" type="mpeg7:unsigned15"/>
      </sequence>
   </sequence>
   </extension>
   </complexContent>
</complexType>
```

8.6.2 Binary representation syntax

Perceptual3DShape {	Number of bits	Mnemonics
NumOfNodes	8	uimsbf
BitsPerAttribute	4	uimsbf
for (i=0; i<(NumOfNodes * NumOfNodes - NumOfNodes)/2; i++) {		
IsConnected[i]	1	bslbf
}		
For (i=0; i< NumOfNodes; i++) {		
Volume[i]	BitsPerAttribute	uimsbf
Center_X[i]	BitsPerAttribute	uimsbf
Center_Y[i]	BitsPerAttribute	uimsbf
Center_Z[i]	BitsPerAttribute	uimsbf
PCA_Axis_1X[i]	BitsPerAttribute	uimsbf
PCA_Axis_1Y[i]	BitsPerAttribute	uimsbf
PCA_Axis_1Z[i]	BitsPerAttribute	uimsbf
PCA_Axis_2X[i]	BitsPerAttribute	uimsbf
PCA_Axis_2Y[i]	BitsPerAttribute	uimsbf
PCA_Axis_2Z[i]	BitsPerAttribute	uimsbf
Max_1[i]	BitsPerAttribute	uimsbf
Max_2[i]	BitsPerAttribute	uimsbf
Max_3[i]	BitsPerAttribute	uimsbf
Convexity[i]	BitsPerAttribute	uimsbf
}		
}		

8.6.3 Descriptor component semantics

NumOfNodes

This field exists only in binary representation and indicates the number of nodes (corresponding to parts in the 3D object). Its minimum value is 1 and maximum is 255.

BitsPerAttribute

The number of bits used for quantization of parameters, Volume, Convexity, Center_X, Center_Y, Center_Z, PCA_Axis_1X, PCA_Axis_1Y, PCA_Axis_1Z, PCA_Axis_2X, PCA_Axis_2Y, PCA_Axis_2Z., Max_1, Max_2 and Max_3. Its minimum value is 1 and maximum is 15. The uniform quantization with a number of steps $2^{\text{BitsPerAttribute}}$ is employed. The parameters residing on the boundaries between bins are assigned to integer value corresponding to the upper bin boundary. That is, all bins are with [min max) style, except that the last bin is [min max]. In more detail, let *n*=BitsPerAttribute, then for a float parameter value *x* in [0, 1] interval, the quantization result *quant(x)* is as follows:

$$quant(x) = \begin{cases} k, \text{ when } \dfrac{k}{2^n} \le x < \dfrac{k+1}{2^n}, \text{ for any integer } 0 \le k < 2^n - 1 \\ 2^n - 1, \text{ when } 1 - \dfrac{1}{2^n} \le x \le 1 \end{cases}$$