
**Information technology — Multimedia
content description interface —**

**Part 3:
Visual**

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

Partie 3: Visuel



PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

© ISO/IEC 2002

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

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.ch
Web www.iso.ch

Printed in Switzerland

Contents

Page

Foreword.....	v
Introduction	vi
1 Scope	1
1.1 Organization of the document.....	1
1.2 Overview of Visual Description Tools	1
2 Terms and Definitions	2
2.1 Default reference axis	2
2.2 DCT coefficients	2
2.3 Data element	3
3 Abbreviations and Symbols	3
3.1 General.....	3
3.2 Abbreviations.....	3
3.3 Arithmetic operators	3
3.4 Logical operators.....	3
3.5 Relational operators	3
3.6 Bitwise operators.....	4
3.7 Conditional operator	4
3.8 Assignment	4
3.9 Mnemonics	4
3.10 Constants	4
3.11 Functions.....	4
4 Conventions	5
4.1 Method of describing the DDL representation syntax.....	5
4.2 Method of describing the binary representation syntax	5
4.3 Method of describing the descriptor semantics	8
5 Basic structures.....	8
5.1 Introduction	8
5.2 Grid layout	8
5.3 Time series	11
5.4 Multiple view	15
5.5 Spatial 2D coordinates	16
5.6 Temporal interpolation.....	23
6 Color.....	29
6.1 Introduction	29
6.2 Color space	29
6.3 Color quantization	33
6.4 Dominant color	35
6.5 Scalable color	37
6.6 Color layout.....	42
6.7 Color structure	50
6.8 GoF/GoP Color.....	56
7 Texture	57
7.1 Introduction	57
7.2 Homogeneous texture.....	57
7.3 Texture browsing.....	61
7.4 Edge histogram.....	63
8 Shape	66
8.1 Introduction	66

8.2	Region shape	66
8.3	Contour shape.....	68
8.4	Shape 3D.....	71
9	Motion	73
9.1	Introduction	73
9.2	Camera motion.....	73
9.3	Motion trajectory.....	81
9.4	Parametric motion	84
9.5	Motion activity.....	87
10	Localization	92
10.1	Introduction	92
10.2	Region locator.....	92
10.3	Spatio-temporal locator	96
11	Others	103
11.1	Introduction	103
11.2	Face recognition	103
Annex A (normative) Basis functions for FaceRecognition		105
A.1 Basis matrix		105
A.2 Mean face		169
Annex B (normative) Binary representation of media time tools.....		171
B.1 Introduction		171
B.2 Binary representation syntax.....		172
B.3 Descriptor components semantics		173
Annex C (informative) Patent statements		174

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 3.

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 part of ISO/IEC 15938 may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 15938-3 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee 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*

Annexes A and B form a normative part of this part of ISO/IEC 15938. Annex C is for information only.

Introduction

This standard, also known as "Multimedia Content Description Interface," provides a standardized set of technologies for describing multimedia content. The standard 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 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
- **Description Definition Language (DDL)** defines Description Schemes, Descriptors, and Datatypes by specifying their syntax, and allows their extension.
- **Systems tools** support delivery of descriptions, multiplexing of descriptions with multimedia content, synchronization, file format, and so forth.

This standard is subdivided into eight 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 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 standard.

Part 7 – Conformance testing: specifies the guidelines and procedures for testing conformance of implementations of the standard.

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

This document contains the visual elements (Descriptors and Description Schemes) that are considered for being part of the standard. All these Descriptive Structures are classified according to the types of visual features they describe. For each Descriptive Structure, there is one corresponding section in this document. The section specifies textual and binary syntax and semantics of the structures.

Information technology — Multimedia content description interface —

Part 3: Visual

1 Scope

1.1 Organization of the document

The structure of this document is as follows. Clauses 2-4 specify the terms, abbreviations, symbols and conventions used throughout the document. Clauses 5-11 contain definitions of the description tools standardized by 15938-3 grouped by the visual features they are associated with, starting with basic structures and containers in Clause 5, through color, texture, shape, motion, localization in Clause 10. Clause 11 contains the remaining, unclassified items.

Each description tool is described by the following subclauses:

- Syntax: Normative DDL specification of the Ds or DSs.
- Binary Syntax: Normative binary representation of the Ds or DSs.
- Semantic: Normative definition of the semantics of all the components of the corresponding D or DS.

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, 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 four color descriptors to represent different aspects of color features: representative colors (DominantColor), color distribution (ScalableColor), spatial distribution of colors (ColorLayout and ColorStructure). It also contains two supporting tools, ColorSpace and ColorQuantization used in DominantColor and an extension of ScalableColor to a group of frames or pictures (GoFGoPColor). 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 Shape3D descriptor provides an intrinsic shape characterization of 3D mesh models.

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 is 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.

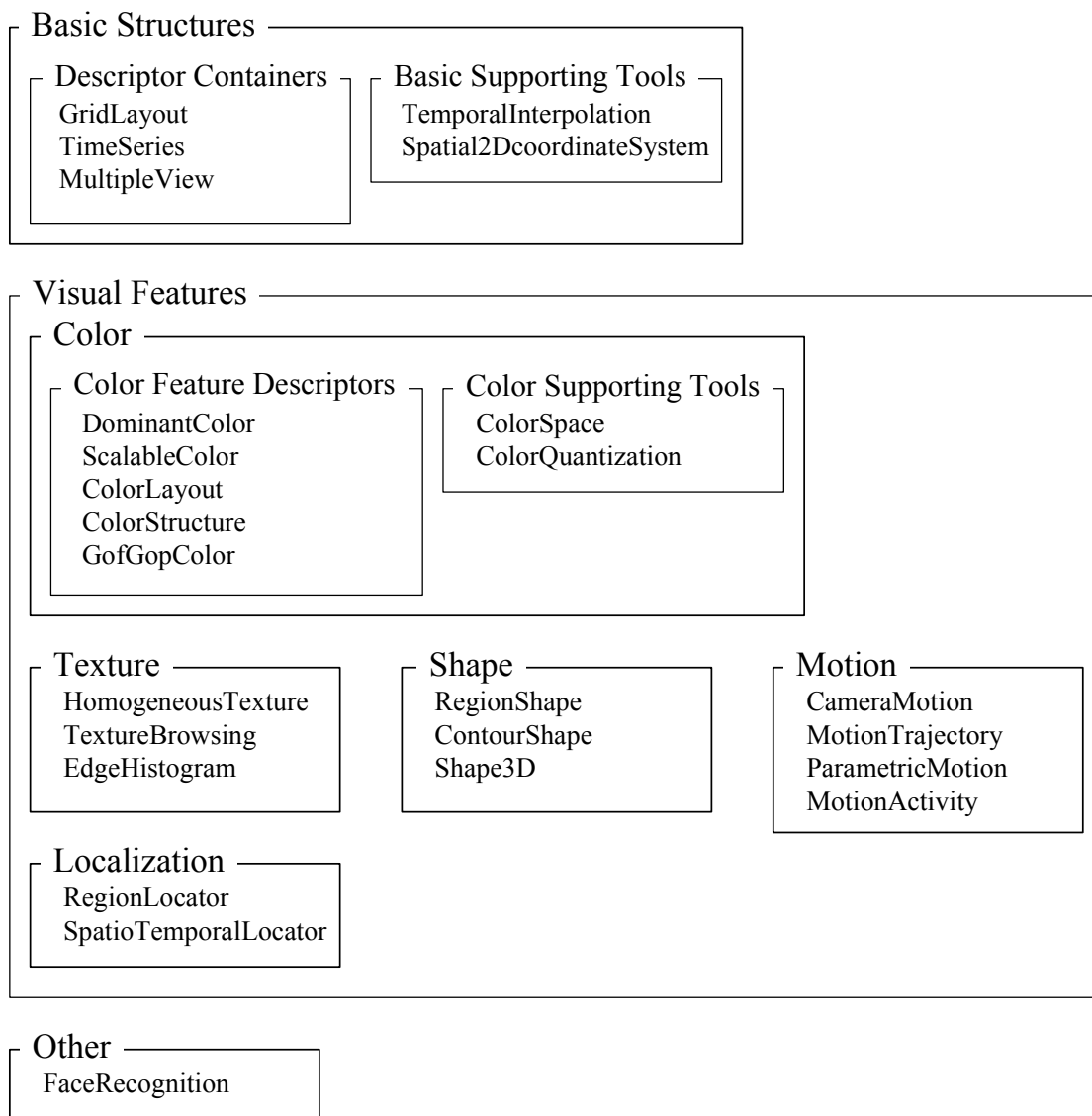


Figure 1 — Overview of Visual Description Tools