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

**Part 9:
Profiles and levels**

**AMENDMENT 1: Extensions to profiles and
levels**

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

Partie 9: Profils et niveaux

AMENDEMENT 1: Extensions aux profils et niveaux



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2012

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.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 1 to ISO/IEC 15938-9:2005 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

Information technology — Multimedia content description interface —

Part 9: Profiles and levels

AMENDMENT 1: Extensions to profiles and levels

After 4.4, add the following subclauses:

4.5 AudioVisual Description Profile (AVDP)

4.5.1 Application Areas

The description tools in this profile can be used to describe the results of various kinds of media analysis with visual and audio low-level features. Consequently the information resulting from media analysis, such as shot/scene detection, face recognition/tracking, speech recognition, copy detection and summarisation, can be used in a wide range of applications such as archive management, news, new services for new media, and many academic projects handling large scale video content, e.g. benchmarking initiatives. The goal is to support establishing an environment where users can exchange evaluation results of media analysis, hence helping the introduction of automatic media analysis tools into many application areas.

4.5.2 Functionality

The main functionality of this profile is describing the results of automatic media analysis with low-, mid- and high-level features for audio-visual content.

The profile supports high-level functionalities such as (but not limited to) the following:

- Provides tools for describing feature extraction tool, version, contributors, and the date/time the tool applied
- Provides tools for describing several results with confidence levels in multiple timelines.
- Provides tools for describing the results of segmentation, text recognition, person identification, format detection, genre detection, keyword extraction, speech recognition, and subject classification.
- Provides tools for describing the results of copy detection and summarization with having several descriptions of audio-visual contents in one document.

4.5.3 Tools in the Profile

This profile is defined with respect to the Version 2 schema as specified in ISO/IEC 15938-10. The namespace of the Version 2 schema providing a basis for this profile is “urn:mpeg:mpeg7:schema:2004.” The following table lists the ISO/IEC 15938 description tools (global elements, global attributes, attribute groups, complexTypes and simpleTypes) and the profile constraints that are part of this profile.

Table AMD 1.1 — Description Profile Tool for AVDP

Global Elements	Name	Constraint
Mpeg7	DescriptionUnit	element excluded
	Description	The document may contain an unlimited number of content descriptions (e.g. for describing a set of contents which contain copies and need to be cross-referenced, or for referencing source segments from a summary). At most one of these descriptions can be of type SummaryDescriptionType.
Global Attributes	Name	Constraint
dim		
Global Attribute Groups	Name	Constraint
audioSpectrumAttributeGrp		
mediaTimePropertyGrp	mediaTimeBase	The XPath shall not contain axis specifiers, i.e. it excludes elements [5] and [6] from the datatype definition in Annex A.1 of ISO/IEC 15938-5. On the root a/v segment, mediaTimeBase may only point to a MediaLocator element in the same description, i.e. the value is limited to be one of • MediaLocator • MediaInformation/MediaProfile/MediaInstance[i]/MediaLocator where i is integer, and can be omitted if only one profile/instance exists.
referenceGrp	idref	attribute excluded
	href	attribute excluded
	xpath	ReferenceType is only used in Availability/InstanceRef. The value of the xpath attribute shall be “../..../MediaInformation/MediaProfile/MediaInstance[i]” where i is integer, and can be omitted if only one profile/instance exists.

Complex Types	Name	Constraint
AdvancedFaceRecognitionType		
AgentType	Icon	element excluded
AudioBPMTType		
AudioDSType		
AudioDType		
AudioFundamentalFrequencyType		
AudioHarmonicityType		
AudioLLDScalarType		
AudioLLDVectorType		
AudioPowerType		
AudioSegmentTemporalDecompositionType	AudioSegmentRef	element excluded
AudioSegmentType	MediaTime	minOccurs = "1"
	TemporalMask	element excluded
	MediaSourceDecomposition	element excluded
AudioSignalQualityType		
AudioSignatureType		
AudioSpectrumBasisType		
AudioSpectrumCentroidType		
AudioSpectrumEnvelopeType		
AudioSpectrumFlatnessType		
AudioSpectrumProjectionType		
AudioSpectrumSpreadType		

AudioSummaryComponentType	Title	element excluded ^{a)}
	AudioSourceID	minOccurs = "1"; AudioSourceID is defined to reference MediaInformation/MediaProfile/MediaInstance[i]/InstanceIdentifier, where i is integer, of the root content segment of a content description.
	AudioSourceLocator	minOccurs = "1"
	ComponentSourceTime	element excluded
	SoundLocator	element excluded
	SyncTime	element excluded
AudioTempoType		
AudioVisualSegmentMediaSourceDecompositionType	AudioVisualSegment	element excluded
	AudioVisualSegmentRef	element excluded
	VideoSegmentRef	element excluded
	StillRegionRef	element excluded
	AudioSegmentRef	element excluded
AudioVisualSegmentTemporalDecompositionType	AudioVisualSegmentRef	element excluded
	AudioVisualRegion	element excluded
	AudioVisualRegionRef	element excluded
AudioVisualSegmentType	MediaTime	minOccurs = "1"
	TemporalMask	element excluded
	SpatialDecomposition	element excluded
	SpatioTemporalDecomposition	element excluded
AudioVisualType		
AudioWaveformType		
AvailabilityType	Financial	element excluded
	Rights	element excluded
BackgroundNoiseLevelType		
BalanceType		
BandwidthType		

BoxListType		
CameraMotionSegmentType		
CameraMotionType		
ClassificationType	Release	element excluded
	Target	element excluded
	ParentalGuidance	element excluded
	MediaReview	element excluded
ColorLayoutType		
ColorQuantizationType		
ColorSamplingType		
ColorSpaceType		
ColorStructureType		
ColorTemperatureType		
CompleteDescriptionType	DescriptionMetadata	element excluded
	Relationships	element excluded
	OrderingKey	element excluded
ConfusionCountType		
ContentAbstractionType		
ContentDescriptionType		
ContentEntityType	MultimediaContent	maxOccurs = "1"
ContourShapeType		
ControlledTermUseType		
CreationInformationType		
CreationToolType		
CreationType	TitleMedia	may only be used on the root a/v segment
CreatorType		
CrossChannelCorrelationType		

DSType	Header	The following headers may be used: <i>Description Metadata</i> • A complete DescriptionMetadata DS is required on each top level temporal decomposition, and optional on the root a/v segment. • A description metadata element that only specifies the Confidence of a description scheme, may be used on any segment whenever the description scheme does not provide any other means to specify the confidence of the description. It is assumed that the confidence levels are not normalised across tools and thus they are not directly comparable. The Confidence shall not occur in description metadata headers in summaries. <i>Headers from part 3 and 4 types</i> may be used on any segment. The use of Header is prohibited on all other elements.
	id	IDs shall be used on the following elements: • All segments • Top-level temporal decompositions IDs may be used on any other description scheme.
	timePropertyGrp	attribute group excluded
	mediaTimePropertyGrp	mediaTimePropertyGrp shall be used on the root a/v segment, and allowed only on TemporalDecompositionType, SpatioTemporalDecomposition-Type SummarizationType and SummaryType. If SummarizationType has mediaTimepropertyGrp, each of the SummaryType elements may have mediaTimePropertyGrp, otherwise each of the element shall have mediaTimePropertyGrp.
DType		

DcOffsetType		
DescriptionMetadataType	Package	element excluded
	*	At least one element shall be present.
DescriptionProfileType		
DisseminationType		
DominantColorType		
DoubleDiagonalMatrixType		
DoubleMatrixType		
EdgeHistogramType		
ElectronicAddressType		
ErrorEventType		
ExtendedLanguageType		
FaceRecognitionType		
FigureTrajectoryType		
FloatDiagonalMatrixType		
FloatMatrixType		
FocusOfExpansionType		
FractionalPresenceType		
GeographicPointType		
GoFGoPColorType		
GofGopFeatureType		
GridLayoutType		
HarmonicInstrumentTimbreType		
HarmonicSpectralCentroidType		
HarmonicSpectralDeviationType		
HarmonicSpectralSpreadType		
HarmonicSpectralVariationType		
HeaderType		

HierarchicalSummaryType	SummaryThemeList	element excluded
	SummarySegmentGroupRef	element excluded
	components	attribute excluded
	hierarchy	fixed="dependent"
HomogeneousTextureType		
IlluminationInvariantColorType		
ImageLocatorType	MediaTimePoint	element excluded
	MediaRelTimePoint	element excluded
	BytePosition	element excluded
ImageTextType		
InlineTermDefinitionType	Name	(See footnote) ^{b)}
	Definition	element excluded
	Term	element excluded
InstrumentTimbreType		
IntegerDiagonalMatrixType		
IntegerMatrixType		
IrregularVisualTimeSeriesType		
KeyType		
KeywordAnnotationType		
LexiconType		
LogAttackTimeType		
MediaAgentType	AgentRef	element excluded
MediaFormatType	Medium	element excluded
	System	element excluded
	Bandwidth	element excluded
	TargetChannelBitRate	element excluded
	ScalableCoding	element excluded
	Emphasis	element excluded
	Presentation	element excluded
	SceneCodingFormat	element excluded
	GraphicsCodingFormat	element excluded
	OtherCodingFormat	element excluded

MediaIdentificationType	AudioDomain	element excluded
	VideoDomain	element excluded
	ImageDomain	element excluded
MediaIncrDurationType	mediaTimeUnit	attribute excluded The value of the mediaTimeUnit attribute on the closest parent TemporalDecomposition or SpatiotemporalDecomposition element applies. If the attribute is not present on any of the parent TemporalDecomposition or SpatiotemporalDecomposition elements, the value of the mediaTimeUnit attribute of the root a/v segment shall be used.
MediaInformationType	MediaProfile	maxOccurs = "1"
MediaInstanceType	LocationDescription	element excluded
MediaLocatorType	InlineMedia	element excluded
MediaProfileType	MediaTranscodingHints	element excluded
	master	attribute excluded
MediaQualityType	RatingInformationLocator	element excluded
MediaRelIncrTimePointType	mediaTimeUnit	attribute excluded The value of the mediaTimeUnit attribute on the closest parent TemporalDecomposition or SpatiotemporalDecomposition element applies. If the attribute is not present on any of the parent TemporalDecomposition or SpatiotemporalDecomposition elements, the value of the mediaTimeUnit attribute of the root a/v segment shall be used.
	mediaTimeBase	attribute excluded The value of the mediaTimeBase attribute on the closest parent TemporalDecomposition or SpatiotemporalDecomposition element applies. If the attribute is not present on any of the parent TemporalDecomposition or SpatiotemporalDecomposition elements, the value of the mediaTimeBase attribute of the root a/v segment shall be used.

MediaSourceSegmentDecompositionType		
MediaTimeType	MediaTimePoint	element excluded
	MediaRelTimePoint	element excluded
	MediaDuration	element excluded
MelodyContourType		
MelodySequenceType		
MelodyType		
MeterType		
MinusOneToOneMatrixType		
MixtureAmountOfMotionType		
MixtureCameraMotionSegmentType		
MotionActivityType		
MotionTrajectoryType		
MovingRegionFeatureType		
MovingRegionSpatioTemporalDecompositionType	MovingRegionRef	element excluded
	StillRegionRef	element excluded
MovingRegionType	SpatioTemporalMask	element excluded
	GofGopFeature	element excluded
	MultipleView	element excluded
	SpatialDecomposition	element excluded
	SpatioTemporalLocator	minOccurs = "1"
	TemporalDecomposition	element excluded
	MediaSourceDecomposition	element excluded
Mpeg7BaseType		
Mpeg7Type	DescriptionProfile	minOccurs = "1"
	DescriptionMetadata	minOccurs = "1"
	timePropertyGrp	attribute group excluded
	mediaTimePropertyGrp	attribute group excluded
MultimediaContentType		

NameComponentType		
NonMixtureAmountOfMotionType		
NonMixtureCameraMotionSegmentType		
OrganizationType	ContactRef	element excluded
	Jurisdiction	element excluded
	JurisdictionRef	element excluded
	AddressRef	element excluded
ParameterTrajectoryType		
ParametricMotionType		
PercussiveInstrumentTimbreType		
PersonGroupType	MemberRef	element excluded
	Jurisdiction	element excluded
	JurisdictionRef	element excluded
	AddressRef	element excluded
PersonNameType		
PersonType	OrganizationRef	element excluded
	PersonGroupRef	element excluded
	AddressRef	element excluded
PhoneLexiconType		
PlaceType		
ProbabilityMatrixType		
RatingType		
ReferenceType		
RegionLocatorType		
RegionShapeType		
RegularVisualTimeSeriesType		

RelatedMaterialType	DisseminationFormat	element excluded
	MediaInformation	element excluded
	MediaInformationRef	element excluded
	CreationInformation	element excluded
	CreationInformationRef	element excluded
	UsageInformation	element excluded
	UsageInformationRef	element excluded
RelationType	type	use = "required"
	source	attribute removed ^{c)}
	strength	Note: It is assumed that the strength levels are not normalised across tools and thus they are not directly comparable.
	target	restricted length of list to 1 ^{d)}
RelativeDelayType		
RightsType		
ScalableColorType		
ScalableSeriesType		
SegmentDecompositionType	criteria	restricted to URI pattern ^{e)} criteria of the decomposition deeper than 2 level shall not use the same value of criteria previously used in their ancestor decomposition level, i.e. level 1 and 2. ^{f)}
SegmentType	MediaInformation	MediaInformation shall be used on the root a/v segment and may not be used on any other segment
	MediaInformationRef	element excluded
	MediaLocator	MediaLocator can be optionally used on key frame segments to refer to extracted key frame images
	StructuralUnit	minOccurs = "1" StructuralUnit of the segment deeper than 2 level shall not use the same value of StructuralUnit previously used in their ancestor segment level, i.e. level 1 and 2. ^{g)}
	CreationInformationRef	element excluded