Multimedia Content Adaptation

José M. Martínez

JoseM.Martinez@uam.es
Grupo de Tratamiento de Imágenes, Escuela Politécnica Superior
Universidad Autónoma de Madrid
http://www-gti.ii.uam.es

Index

Credits

Part of this presentation is based on the research of some members of the Image Processing Group (Grupo de Tratamiento de Imágenes) at Universidad Autónoma de Madrid

- ★ Fernando Barreiro
- ★ Jesús Bescós
- ★ Luis Herranz
- ★ Fernando López
- ★ José M. Martínez
- ★ Fabricio Tiburzi
- ★ Víctor Valdés
- ★ Also to acknowledge
  - Víctor Fernández-Carbajales
  - Javier Molina
  - And *anonymous models* in the content: GTI@UAM members, Fernando Barreiro's friends, …

Multimedia Content Adaptation: A short overview

Introduction
Content adaptation dimensions
Related Standards
Conclusions
References

Introduction (I): Current Multimedia content delivery scenario

In each session there are different network conditions, terminal capabilities, user (preferences/handicaps)

Different "media spaces":

- Formats, resolutions, bit rate, size, users, …

There is no universal format for all the media spaces.

- It is not expected to converge to a unique universal format.

But media spaces end-to-end interoperability/interworking is required.

- It is needed to transform (adapt) the content among different media spaces.
  - Scalable coding will be an alternative in some (all?) cases.

Variations vs. Universal Multimedia Access

Introduction (II): Variations

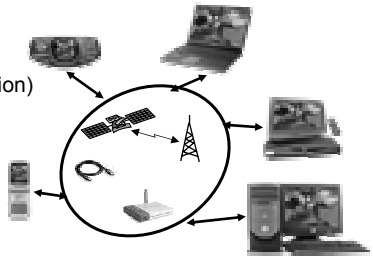
Variations: different versions for each situation (terminal, network, …)

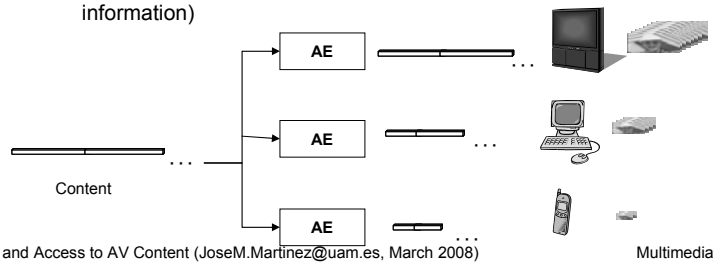
- Generation
  - Previous (*off-line*)
    - Huge amount of storage resources
    - Instantaneous availability
  - On demand (*on-line*)
    - Huge amount of processing resources
    - Generation delay
    - The only solution for live content
  - Hybrid
    - Cache
- This was a solution in the "early days"
  - Nowadays there are thousands of (growing) situations [media spaces]

Introduction (III): Universal Multimedia Access

Universal Multimedia Access: Create once, present anywhere, anyhow

- Generic Adaptation
  - Any type of content
  - Any type of terminal (creation and consumption)
  - Any type of network
  - Any type of user and usage environment
- Resource consumption optimization
  - Do not deliver what will not be consumed
  - On demand or hybrid models
  - Optimization of the generation of variations (storage of the required analysis information)





Multimedia Content Adaptation: A short overview

Introduction
Content adaptation dimensions
Related Standards
Conclusions
References

Content adaptation dimensions

Adaptation taking into account …

- The user
- The terminal
- The network
- The environment

Adaptation by means of …

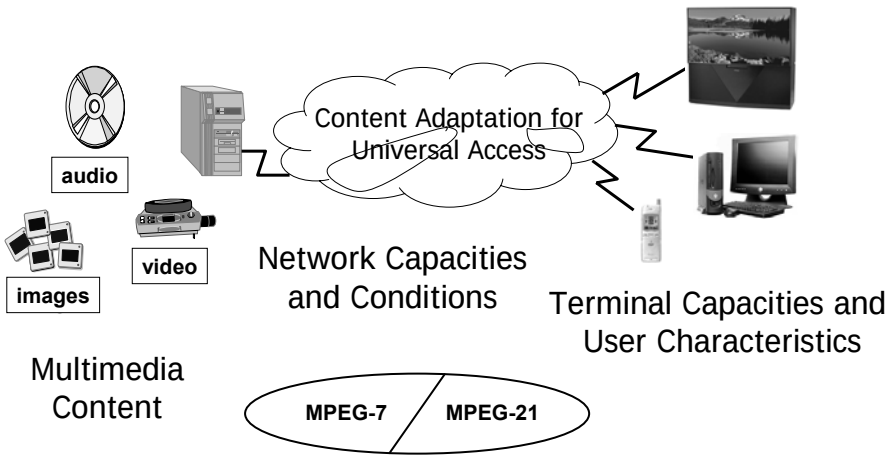
- Transcoding
- Scalable Content (truncation)
- Transmoding and Content Abstraction

Adaptation at

- Signal/Media (media space/format dependent, content independent) level
- Semantic (events, objects, … content dependent) level

Adaptation taking into account … (I)

The need of metadata: There is the need of filling the gap between media spaces/formats and the usage environment: terminal, network, users, environmental conditions, …



Adaptation taking into account … (I)

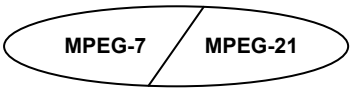
The need of metadata: There is the need of filling the gap between media spaces/formats and the usage environment: terminal, network, users, environmental conditions, … It is required **to provide negotiation and understanding capacities**.

```
<?xml version=''1.0'' encoding=''UTF-8''?>
<MPEG-7 xmlns:mpeg7="urn:mpeg:mpeg7:schema:200">
    <Description xsi:type="ContentEntityType">
     <MultimediaContent xsi:type="VideoType">
      <Video>
       <MediaInformation>
        <MediaProfile master="true">
         <MediaFormat>
          <FileSize>666478608</FileSize>
          <VisualCoding>
           <Format>MPEG-2 Video</Format>
           <Frame height="288" width="352" rate="25"/>
          </VisualCoding>
         </MediaFormat>
        </MediaProfile>
       </MediaInformation>
      </Video>
     </MultimediaContent>
    </Description>
</MPEG-7>
```

```
<?xml version=''1.0'' encoding=''UTF-8''?>
<DIA xmlns:dia=''urn:mpeg:mpeg21:dia:schema:2003''>
    <Description xsi:type="UsageEnvironmentType">
        <Terminal>
            <InputOutput>
                <Display>
                    <Resolution horizontal="176" vertical="144"/>
                </Display>
            </InputOutput>
        </Terminal>
        <Network>
            <NetworkCondition maxCapacity="64000"/>
        </Network>
    </Description>
</DIA>
```

Terminal Capacities and User Characteristics

Multimedia Content



Adaptation taking into account … (II)

User characteristics specify user preferences for some kind of media (video versus images), values of parameters, … as well as handicaps.

- Presentation preferences: Volume, equalization, brightness, contrast, …
- Accessibility characteristics (handicaps): Audition thresholds, colour blindness degree, …

Terminal characteristics specify the capacity and resources of the terminals

- Codec capacities: Media modalities, formats (profile@level), …
- Input/output capacities: Media modalities, esolution, colour depth, …
- Device resources: Battery, free disk space, free RAM, …

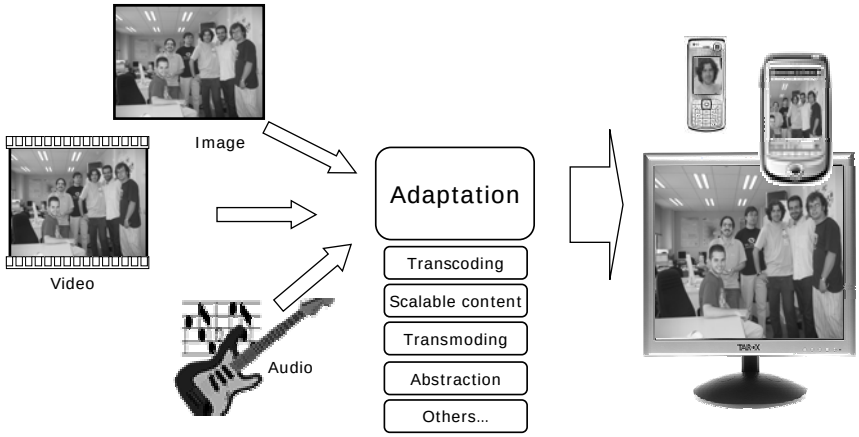
Network characteristics specify the network capacities and condition

- Capacities (static/nominal): Maximum bitrate, guaranteed bitrate, …
- Conditions (dynamic): Available bitrate, delay, error probability, …

Environmental characteristics specify audio-visual conditions of the environment

- Audiovisual conditions: Noise level, noise spectrum, illumination characteristics, …

Adaptation by means of … (I)



Adaptation by means of … (I): Transcoding

Transcoding modalities

- Format change
  - MPEG-2 to MPEG-4, PCM audio to MP3, JPEG to JPEG2000, …
- Scaling
  - Spatial resolution, temporal resolution, number of colours, …
- Quality
  - Bits per sample, “Frequency quality” (subbands, AC coefficients, …), ...

Transcoding approaches

- Full decoding and encoding (computationally inefficient)
- Partial decoding and encoding (may produce artefacts and additional distortion due to approximations)

Transcoding is computationally inefficient but the only solution in most cases

- Storage of transcoding hints (analyze once, transcode several times)

Adaptation by means of … (II): Scalable content

Scalable Coding Formats
- A *single* bitstream contains *multiple (hierarchical)* versions of the same content
- Allows to adapt a content by bitstream "truncation" instead of decoding and reencoding

Scalability types
- Spatial: resolution (less pixels)
- Temporal: frame rate (less frames per second)
- Quality: SNR (subbands)

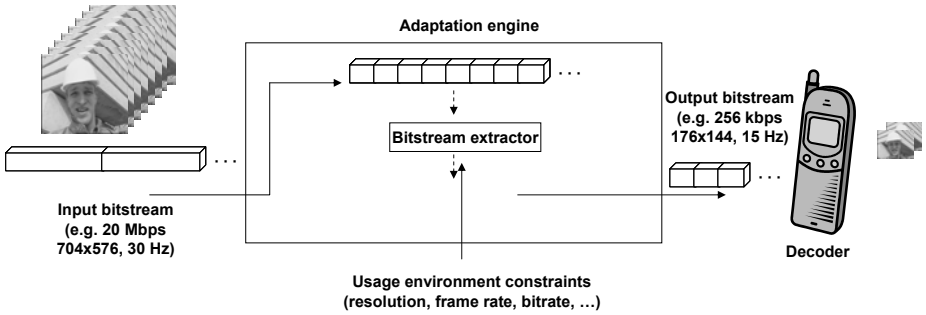
Granularity
- Levels/stages
- *Fine Grain Scalability*
- *Real scalability* (the utopic objective)

Compression efficiency
- The challenge in Video

There are different standard formats including profiles for scalable coding (don't be confused by progressive coding)
- MPEG-2, MPEG-4, H.264
- JPEG 2000

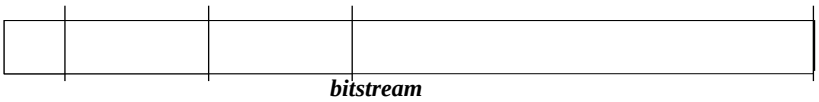
Adaptation by means of … (V): Scalable content - Scalable adaptation

Adaptation
- Simply bitstream extraction
  - Very efficient



Adaptation by means of … (III): Scalable content – Spatial-SNR scalability

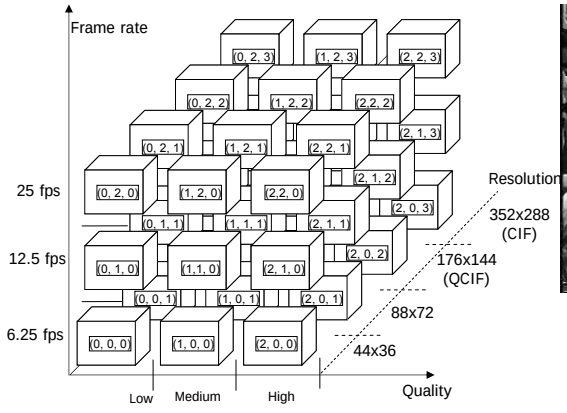
Scaled for slides (full res.)



Adaptation by means of … (VI): Scalable content – Spatial-SNR-Temporal scalability

Cube structure

- Adaptation unit: Group of frames (GoF)

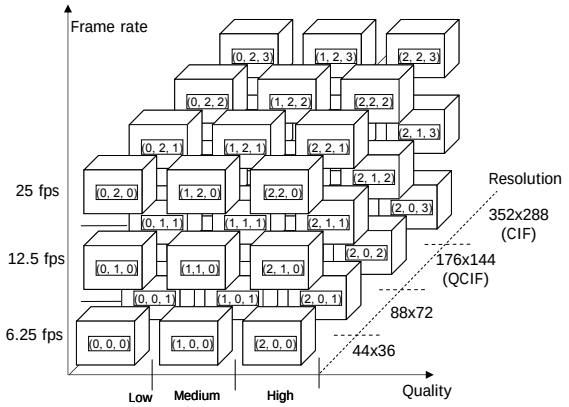


352x288, 25fps, High bitrate

Adaptation by means of … (VII): Scalable content – Spatial-SNR-Temporal scalability

Cube structure

- Adaptation unit: Group of frames (GoF)
- Spatial Resolution adaptation

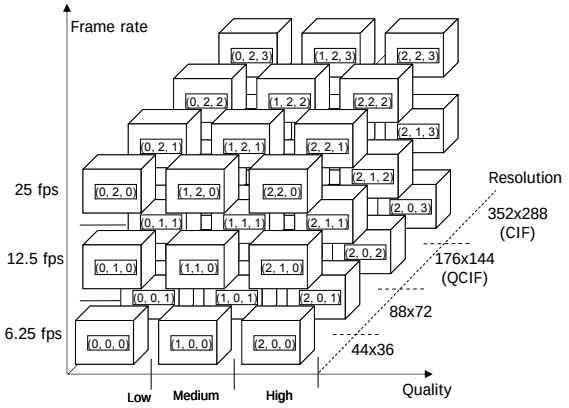


88x72, 25fps, High bitrate

Adaptation by means of … (VIII): Scalable content – Spatial-SNR-Temporal scalability

Cube structure

- Adaptation unit: Group of frames (GoF)
- Spatial resolution + Quality resolution adaptation

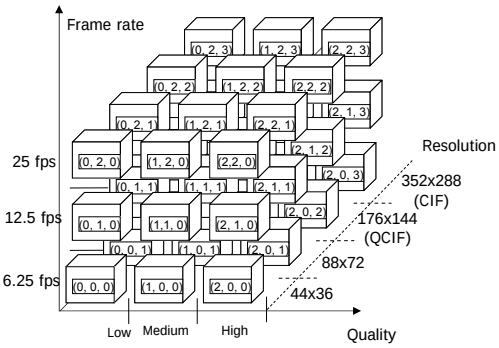


88x72, 25fps, **Low bitrate**

Adaptation by means of … (IX): Scalable content – Spatial-SNR-Temporal scalability

Cube structure

- GoF length 8 frames
- Temporal resolution adaptation

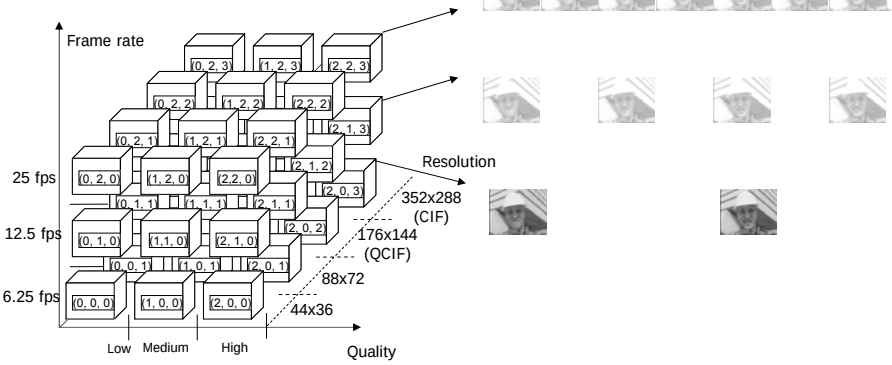


352x288, **25fps**, High bitrate

Adaptation by means of … (X): Scalable content – Spatial-SNR-Temporal scalability

Cube structure

- GoF length 8 frames
- Temporal resolution adaptation

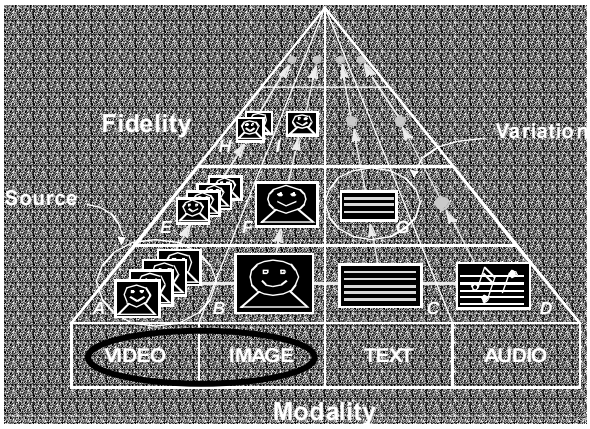


352x288, **6.25fps**, High bitrate

Adaptation by means of … (XI): Transmoding

Transmoding (Info-pyramid)

- Change of modality of the content



Adaptation at … (I): Semantic based adaptation

The adaptation is not at frame (soundtrack) level, but driven by the content present in the media

- Usually in adaptation some information is lost, therefore the idea is to try to preserve the part with more "relevant" information
- Requires content analysis (off-line or on-line)
  - Regions of Interest
    - Spatial: segmentation in "homogeneous" regions: Can be spatio-temporal ones (segmentation + tracking)
    - Rectangular, Polygonal, Generic, …
    - Automatic/supervised extraction: object recognition (by shape, texture, colour, …)
  - Abstraction (Summaries)
    - Temporal: segmentation in shots/sequences
    - Sequential and hierarchical (representation/browsing level)
    - Automatic/supervised extraction: event recognition (based in objects, relations, motion, rules, context restrictions, …)

Adaptation at … (II): ROI based semantic spatial adaptation

Blind resolution adaptation

Blind cropping

Semantic resolution adaptation

Adaptation at … (III): ROI based semantic SNR adaptation

Blind adaptation vs. Semantic adaptation

Adaptation at … (IV): Video abstraction

Need for efficient browsing and retrieval

- Typical solution in multimedia databases: each result is represented by textual information and/or a keyframe (e.g. YouTube)

Video abstraction techniques

- Give a simplified representation quick glance of what happens in the video content
- Should preserve those parts more "semantically" relevant and discard those less informative
  - Requires analysis of the content
  - Otherwise uniform sampling of video

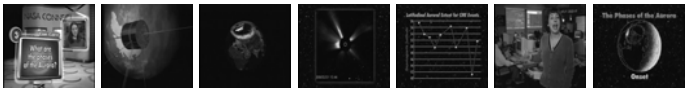
    Some studies reveal that for some applications may be enough

(minimal) Video abstraction taxonomy

- Video summarization: (usually) keyframe representations
  - Storyboards, slide shows, video posters
- Video skimming: shot (or other kind of video unit) representations
  - Summary oriented or Highlights oriented

Adaptation at … (V): Video abstraction (II)

Video summarization example:
Image storyboard



Video skimming example:
Fast playback

Multimedia Content Adaptation: A short overview

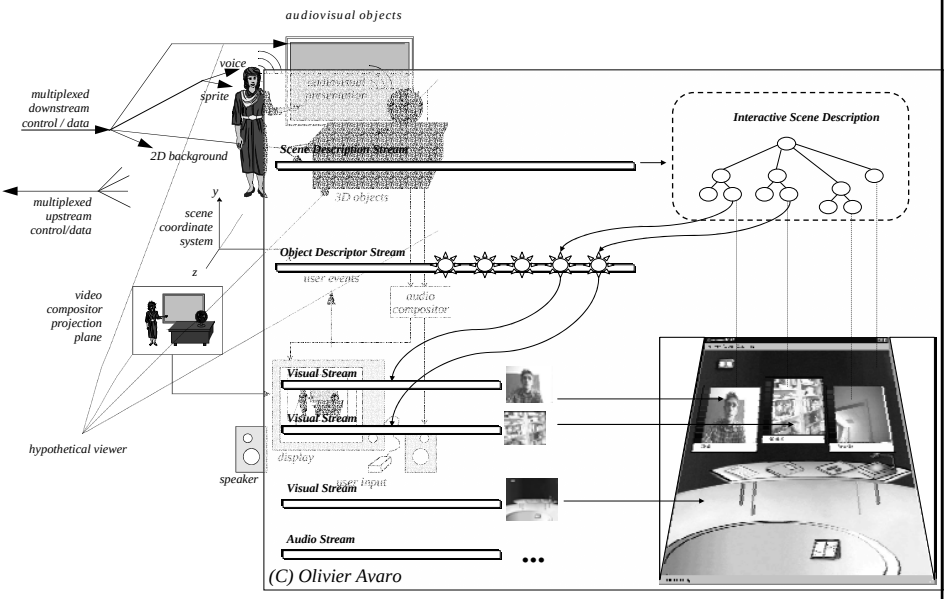
Introduction
Content adaptation dimensions
Related Standards
Conclusions
References

Related Standards (I)

MPEG

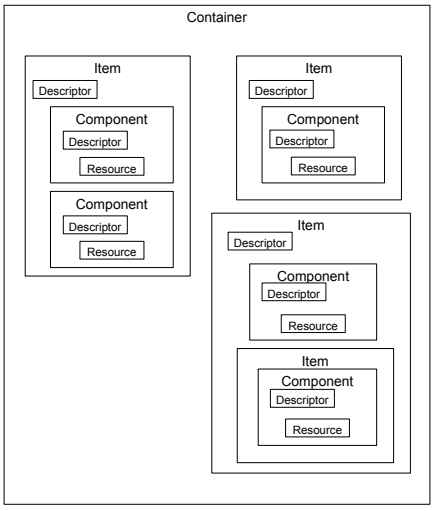
- They are not the only ones dealing with adaptation, but they conform the more complete family, covering all the UMA related aspects
- System level Description format (AV scene, *Digital Item*)
  - MPEG-4 Systems (Selection of elementary streams from an AV scene)
  - MPEG-21 DID (Selection of components and items from a container of resources)
- Coding formats
  - MPEG-1/2/4 Video and Audio (Content scalability)
- Metadata
  - MPEG-7 (Content Description)
  - MPEG-21 DIA (Usage environment description, adaptation tools)

Related Standards (II): MPEG-4 adaptation at system level



Related Standards (III): MPEG-21 Digital Items

MPEG-21: Part 2 – Digital Item Declaration



Related Standards (IV): MPEG-7 y MPEG-21 - Metadata for adaptation

MPEG-7

- Content Description
  - Description of multimedia formats, archival information, …
  - Description of variations and summaries
  - Description of transcoding hints
  - Description of semantic content

MPEG-21: Part 7 – Digital Item Adaptation

- Description of terminal, network, user and environmental conditions
- Tools for multimedia resources adaptation
  - XML representation of the structure of scalable formats for format independent adaptation (BSD and gBSD)
  - Relation ships among usage environment constraints, feasible adaptations and quality of results at QoS level (AdaptationQoS)
  - Tools for metadata adaptation

For a short overview see Annex: MPEG-7 and MPEG-21 metadata for content adaptation

Multimedia Content Adaptation: A short overview

Introduction
Content adaptation dimensions
Related Standards
Conclusions
References

Conclusions

Multimedia content adaptation is a key factor in the current (and future) multimedia communications scenario: multimedia-multiuser-multiterminal-multinetwork

Adaptation must take into account different aspects: user, terminal, network, environment, …, and rights (copyrights and author rights)

Adaptation can be performed by transcoding, truncation of scalable content and transmoding

- In order to enhance the user' experience (Quality of Experience versus QoS) adaptation should **take semantics into account**.
- The best solution depends on the application and usage environment

There are already international standards supporting content adaptation within the framework of Universal Multimedia Access.

Multimedia Content Adaptation: A short overview

Introduction
Content adaptation dimensions
Related Standards
Conclusions
References

References

- A. Vetro, C. Christopoulos, T. Ebrahimi (eds.), "Universal Multimedia Access (special issue)", *IEEE Signal Processing Magazine,* 20 (2), March 2003.
- F. Pereira, I.S. Burnett, S.-F. Chang, "Special Issue on Multimedia Adaptation", *Image Communication*, 18(8), September 2003.
- G. Singh (ed.), "Content repurposing (special issue)", *IEEE Multimedia*, 11 (1), Jan-Feb 2004.
- R. Cucchiara, A del Bimbo, "Special Issue on Video Segmentation for Semantic Annotation and Transcoding", *Multimedia Tools and Applications*, 26(3), August 2005.
- F.Pereira, P.v.Beek, A.C.Kot, J. Ostermann, "Special issue on analysis and understanding for video adaptation", *IEEE Trans. on Circuits and Systems for Video Technology*, 15(10), October 2005.
- *Plus hundreds of papers on:*
  - *transcoding, scalable coding, adaptation, summarization, repurposing, semantic adaptation, content-based adaptation, multimedia analysis for content undertanding and adaptation, …*

MPEG

- MPEG Home Page, http://www.chiariglione.org/mpeg/
- "The MPEG-4 Book", F. Pereira, T. Ebrahimi (eds.), Prentice-Hall, 2002.
- "Introduction to MPEG-7", B.S. Manjunath, P. Salembier, and T. Sikora (eds.), John Wiley & Sons, 2002.
- "The MPEG-21 Book", I.S. Burnett, F. Pereira, R.v.d. Walle, R. Koenen (eds.), John Wiley & Sons, 2006.

Index

Multimedia Content Adaptation: some experiences

The CAIN adaptation engine (CAIN)

Semantic adaptation of images (SAI)

Image2Video transmoding (I2V)

On-line semantic video abstraction (OLSVA)

Semantic adaptation of scalable video (SASV)

…

CAIN (I): Introduction

Content Adaptation Engines are software modules in charge of managing the adaptation of content to the different UAM scenarios

There are different approaches towards Content Adaptation within the UMA framework

- Transcoding
- Transmoding
- Scalable coding
- Abstraction (Summarization)
- Semantic driven adaptation (vs. blind adaptation)
- …

CAIN stands for Content Adaptation Integrator

- Content adaptation engine targeted to the integration of *different but complementary* content adaptation approaches
- Metadata-driven (Content and Context Usage descriptions)

CAIN (II): Objectives

CAIN aims to select the most *appropriate* available Content Adaptation Tool (CAT)

- Efficiency
- Computational cost
- Quality (of experience) of adapted media

based on the constraints imposed by

- Media format
- Usage Environment
  - Terminal
  - Network
  - Environmental conditions
  - …
- User Preferences
  - (May be) different for different usage environment conditions

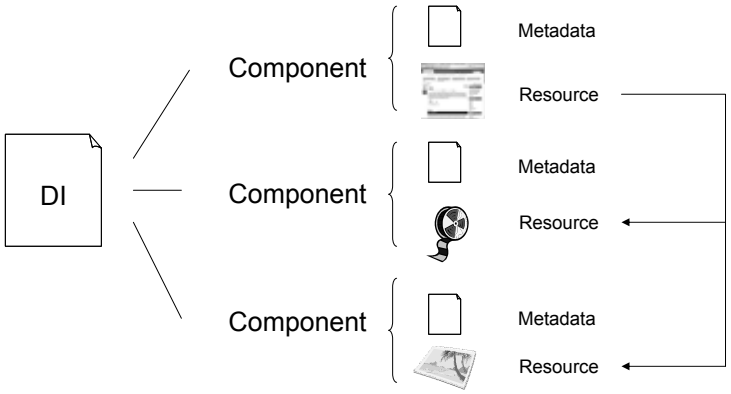
CAIN (III): Functionalities

CAIN is a media processing node in charge of:

- Selecting the appropriate media adaptation parameters and modality based on the available metadata
  - Media description
  - Usage environment description
  - Low-level and Semantic annotations
- Selecting the appropriate Content Adaptation Tool (CAT)
- Configuring and managing the selected CAT to deliver adapted content
  - To a content repository
  - To a streaming service

CAIN (IV): Digital Item adaptation

CAIN is an MPEG-21 compliant adaptation engine in the sense that it receives and produces a Digital Item (DI)



CAIN (V): CAT, CAT Caps, conversions and adaptations

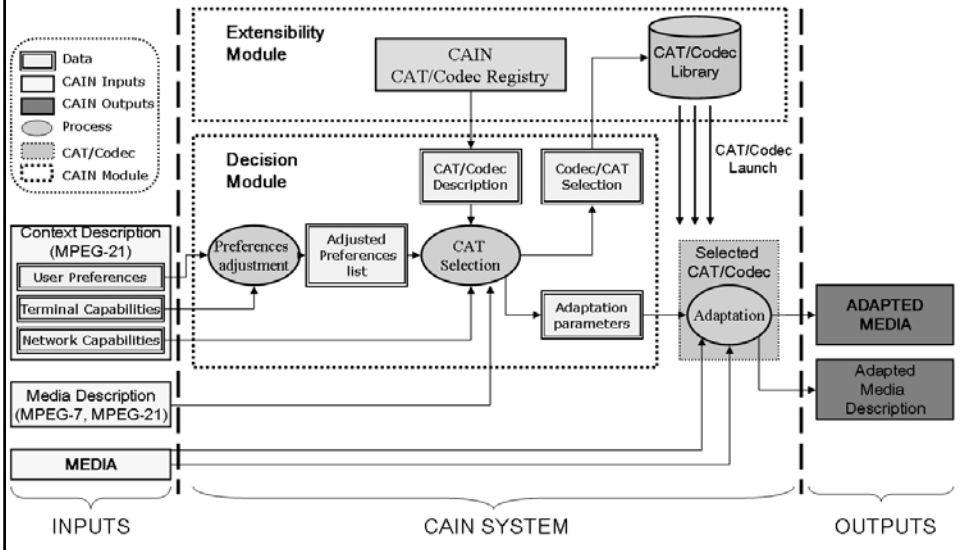
CAT. A software module capable of performing different kinds of conversions over one specific *Component* of the DI.

CAT Capabilities. A document where the conversions that a CAT is capable of performing are annotated.

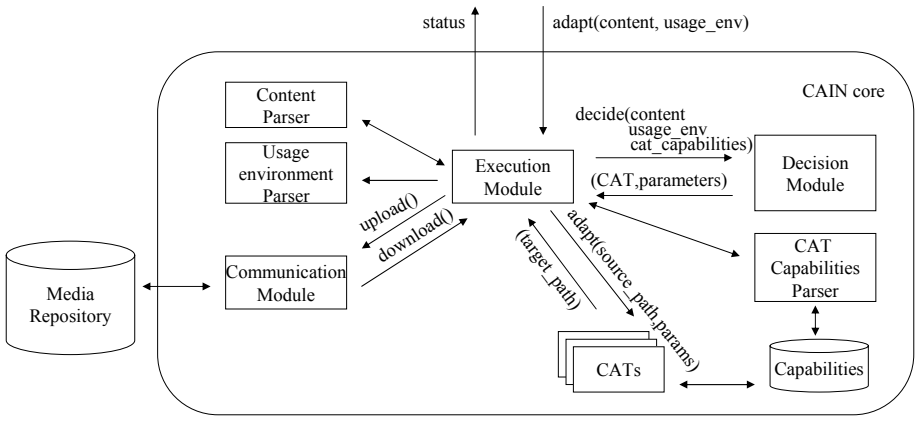
Conversion. Specific action and parameters over one *Component* of the DI.

Adaptation. The most typical conversion where the *Component* is modified to fulfil the usage environment. May imply one or more conversions.

CAIN (VI): Architecture



CAIN (VII): Adaptation Process Overview



CAIN (VIII): Current Content Adaptation Tools

Transcoder CAT
- Supports transcoding of
  - Video: MPEG-1/2/4, H.264, RAW
  - Audio: MPEG-1 layer 2 and layer-3, MPEG-4 AAC, AMR, RAW
  - Image: JPEG, GIF, PNG, RAW
- Based on ffmpeg

Scalable Image CAT
- Supports JPEG2000 content adaptation
- Based MPEG-21 DIA's BSD

Semantic Image CAT
- Supports Semantic image adaptation:
  - JPEG2000 (SNR adaptation)
  - GIF, JPEG, JPEG2000, … (spatial adaptation)
- Currently based on people presence

Image2Video CAT
- Supports Image tranmoding to a video
- Currently based on people presence

Real-time Content-based CAT
- Supports MPEG-1/2 video
- Uses information from the MPEG compressed domain
- Currently temporal summarization and spatial adaptation by cropping of region of interest (motion)

CAIN (IX): Description processing

Content descriptions based on MPEG-7 MDS and MPEG-21 DIA BSD

Adaptation type	Content description
Media adaptation	MPEG-7 Media Description (MediaInformation, Transcoding Hintes)
	MPEG-7 Variations and Summaries
Semantic adaptation	Regions of interest with importance (annotated by users or automatically)
Transmoding to text	MPEG-7 textual tools: Keywords, textual annotations, Spoken Content
Bitstream adaptation	MPEG-21 BSD/gBSD. Preferably gBSD to allow semantic bitstream adaptation

Context descriptions based on MPEG-21 DIA Usage Environment tools

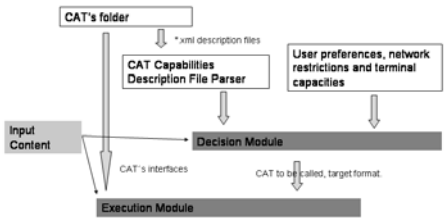
User Description Tools		
Usage Preferences	**Presentation Preferences**	**Accessibility Characteristics**
•Media Format: content, bit rate, visual coding (format, frame height, frame width, frame aspect ratio and frame rate), audio coding (format, audio channels, sample rate, bits per sample).	•AudioPresentationPreferences: Volume, output device, balance. •DisplayPresentationPreferences: Color temperature, brightness, saturation, contrast. •ConversionPreferences: Media type conversion preferences and priorities. •PresentationPriorityPreferences: Modality (audio, video...) priorities.	•No tools from this group.

Terminal Capabilities Tools			
Codec Capabilities	**Display Capabilities**	**Audio Output Capabilities**	**Storage Characteristics**
•Audio, video and image coding/decoding supported formats.	•Supported display modes (resolution, refresh rate), screen size, color bit depth.	•Supported audio modes (sampling frequency, bits per sample), low frequency, high frequency, number of channels…	•Input transfer rate, output transfer rate, size, writable.

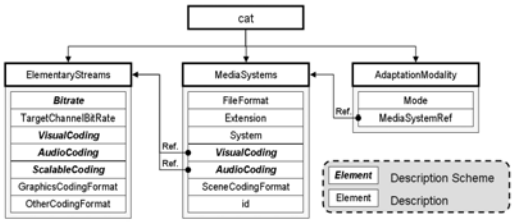
Network Characteristics Tools
•Network Capability
•Maximum capacity and minimum guaranteed.

CAIN (X): Extensibility in CAIN

★Extensibility allows to dynamically add new CATs



★CAT are described using CAT Capabilities



CAIN (XI): The Decision Module (DM)

The target of the DM is to generate the **feasible sequences of conversions**, and after that, to choose the **optimal sequence of conversions**, that is, the sequence of conversions that fits best with the preferences.

The DM is the module in charge of selecting:

- CAT to execute
- Parameters to execute the CAT

Usage environment is divided in:

- Mandatory usage environment: *Must be fulfilled*
  - Terminal and network constraints
- Desirable usage environment: *Wish list*
  - User preferences

The system is modelled by constraints and we apply constraint matching and optimization for the selection

- First mandatory parameters allow to select candidate CATs and parameters (matching). **Usage environment based decisions**: Search for the feasible sequences of conversions that produce a content that fulfil the constraints of the usage environment (mainly terminal and network).
- Finally desirable parameters allow to select the "best" CAT (optimization). **Preference based decisions**: Search for the optimal sequences of conversions that produce content that fits best with the user and adaptation engine preferences.

CAIN (XII): Conclusions

CAIN aims providing an integrated framework for selecting the most appropriate (from a QoE point of view) content adaptation tools (and approach) based on content and context descriptions

- Transcoding, Scalable content, Real-time content based, transmoding, semantic adaptation, …
- MPEG-7 and MPEG-21 are the selected metadata standards

Results prove that

- Metadata-driven adaptation and selection of different CATs works, supporting flexibility and extensibility
- CAIN can be used for service
  - Prototyping
  - Deployment (customizing the selected CATs and reducing the flexibility)
  - and in a (near –we hope-) future for benchmarking

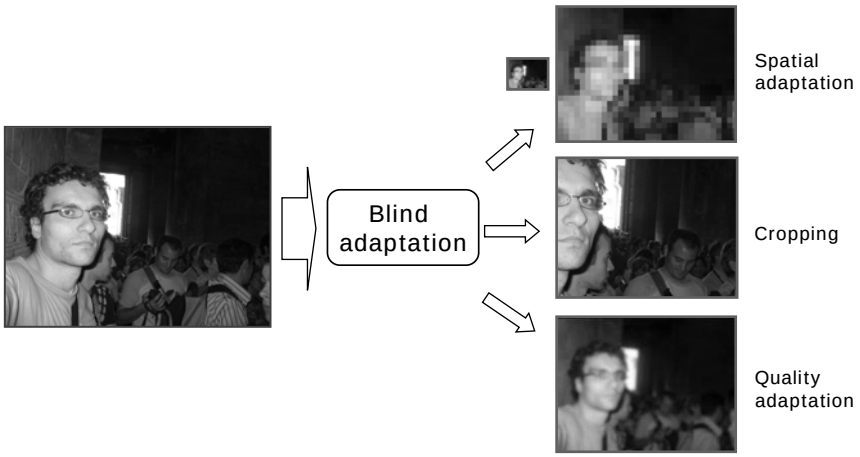
CAIN (XIII): References

- ★ José M. Martínez, Jesús Bescós, Víctor Valdés, Luis Herranz, “Integrating Metadata-driven Content Adaptation Approaches”, en Knowledge-based Media Analysis for Self-adaptative and Agile Multimedia Technology, P. Hobson, E. Izquierdo, Y. Kompatsiaris, N. O’Connor (eds.), EWIMT 2004, Londres, Reino Unido, Noviembre 2004, pp. 18-28
- ★ José M. Martínez, Víctor Valdés, Jesús Bescós, Luis Herranz, “Introducing CAIN: a Metadata-driven Content Adaptation Manager Integrating Heterogeneous Content Adaptation Tools”, en Proceedings of the International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS’2005, Montreux, Suiza, Abril 2005.
- ★ Víctor Valdés, José M. Martínez, “Content Adaptation Tools in the CAIN framework”, en Visual Content Processing and Representation, L. Atzori, D.D. Giusto, R. Leonardi, F. Pereira (eds.), Lecture Notes in Computer Science, Vol. 3893, Springer Verlag, 2006, pp. 9-15
- ★ Víctor Valdés, José M. Martínez, “Content Adaptation Capabilities Description Tool for Supporting Extensibility in the CAIN Framework”, en Multimedia Content Representation, Classification and Security-MCRS2006, B.Günsel, A.K.Jain, A.M. Tekalp, B. Sankur (eds.), Lecture Notes in Computer Science, Vol. 4105, Springer Verlag, 2006, pp. 395-402
- ★ Fernando López, José M. Martínez, Víctor Valdés, “Multimedia Content Adaptation within the CAIN framework via Constraints Satisfaction and Optimization”, Proceedings of the Fourth International Workshop on Adaptative Multimedia Retrieval-AMR06, Ginebra, Suiza, Julio 2006, 17 páginas (CD-ROM).
- ★ Javier Molina, José M. Martínez, Víctor Valdés, Fernando López, “Extensibility of adaptation capabilities in the CAIN content adaptation engine”, Proceedings of the First Internacional Conference on Semantic and Digital Media Technologies, SAMT 2006, Atenas, Grecia, Diciembre 2006.

MCA: some experiences

SAI (I): Introduction

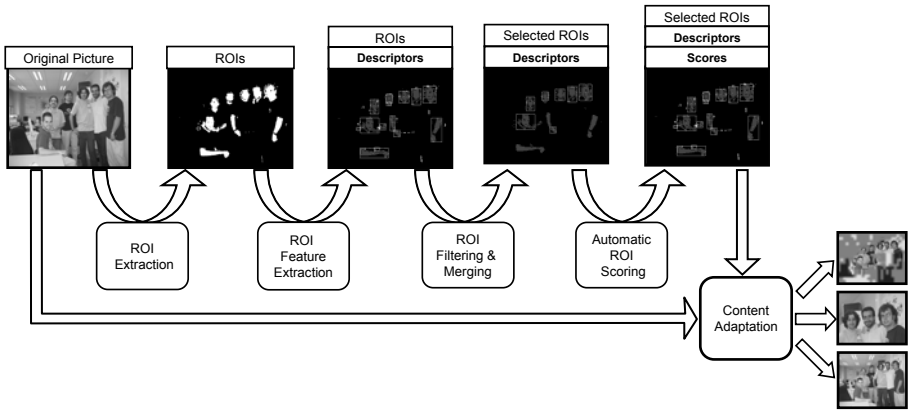
Semantic Adaptation targets to minimize the lost of relevant information when performing image adaptation



SAI (I): Introduction

- ★ Semantic based iamge adaptation depends mainly on the generation of ROIs, which is based on segmentation, a costly process.
- ★ In order to try to minimize the time and resources we propose a ROI management procedure starting with rough segmentation.
- ★ The proposed segmentation is performed applying low complexity algorithms and therefore the process consumes not as much resources as true segmentation ones.
- ★ Although the proposed methods can be general, we present a method based on skin identification.

SAI (III): System Modules

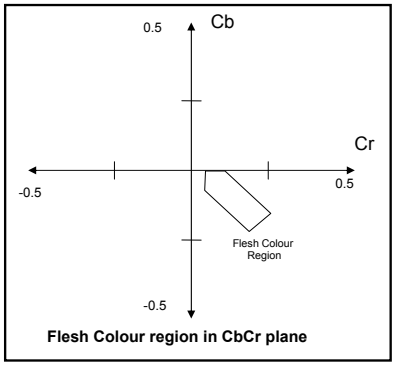


SAI (IV): ROI extraction – Binary segmentation

- ★ The driving semantic feature for image adaptation in this work is ‘people presence’.
- ★ The system is customized based on the assumption of the presence of people in the image and that people is the significant content of the image.
- ★ The selected low-level feature for image segementation is flesh colour.

SAI (V): ROI extraction – Binary segmentation

- ★ The defined colour region in CbCr plane allows flesh colour segmentation with independence of the possible differences between flesh colour intensity and races.
- ★ The decision criterion about which pixels are flesh coloured is taken by geometrically determining if each pixel colour value is inside or outside the selected region.



Flesh Colour region in CbCr plane

SAI (VI): ROI extraction – Binary segmentation

Original Image

Binary Mask

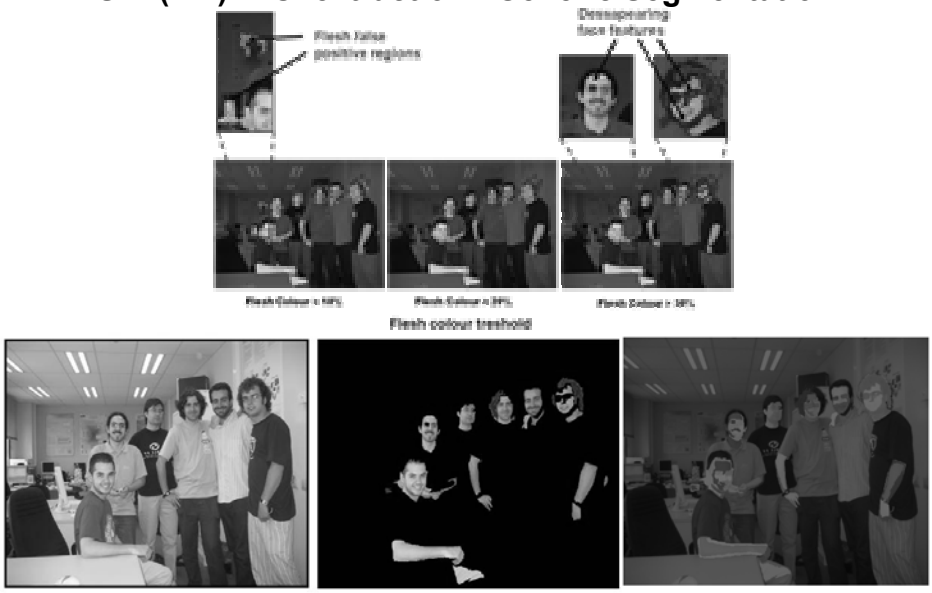
SAI (VII): ROI extraction - Generic Segmentation

Another approach to image segmentation (with drawbacks and advantages) is to perform a generic image segmentation and afterwards perform a selection of regions based on features of the segmented regions.

Good segmentation Under-segmentation Over-segmentation

Results obtained by aceToolbox (Dublin City University)

SAI (VIII): ROI extraction - Generic Segmentation



Original Image **Selected regions (original colour)** **Selected regions (false colour)**

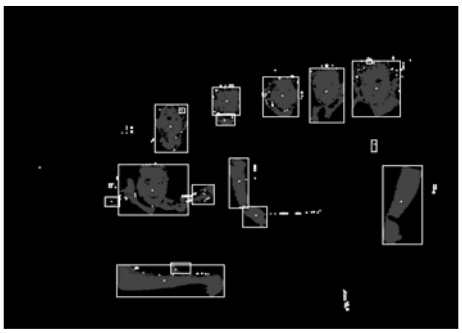
SAI (IX): ROI feature extraction

- ★ The starting point for ROI definition is the segmentation mask indicating, in our case, flesh colour or not.
- ★ The first set of ROIs in the image is created by just creating one ROI for each set of connected group of flesh coloured pixels.
- ★ Each ROI is defined by a list of pixels (position of each pixel) and the bounding box which covers all the pixels in the box.
  - Considering rectangular bounding box aligned with the x and y axis of the image.

SAI (X): ROI feature extraction

The ROI creation process outputs a lot of regions due to the existence of a lot of non significant regions of pixels (either isolated pixels or small regions).

Binary Mask

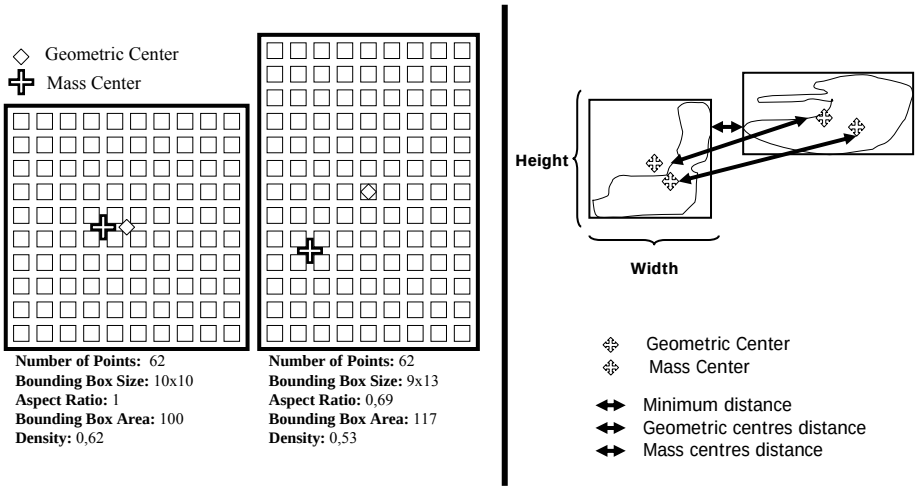


Initial set of ROIs

SAI (XI): ROI feature extraction

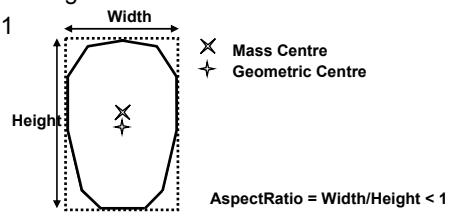
- ★ In order to reduce or even avoid the high number of ROIs number, there is the need of some tools for filtering, merging and ranking the ROIs.
- ★ Besides low-level visual descriptors proposed by MPEG-7, the following additional descriptions has been considered for classificaion of the ROIs for further processing:
  - Number of pixels
  - Bounding box size
  - Bounding box aspect ratio
  - Bounding box area
  - Region geometric centre
  - Region mass centre
  - Geometric-mass centres distance
  - Density: Number of pixels / Bounding box area
  - Vector of distances between ROIs:
    - o Minimum distance
    - o Geometric centres distance
    - o Mass centres distance

SAI (XII): ROI feature extraction



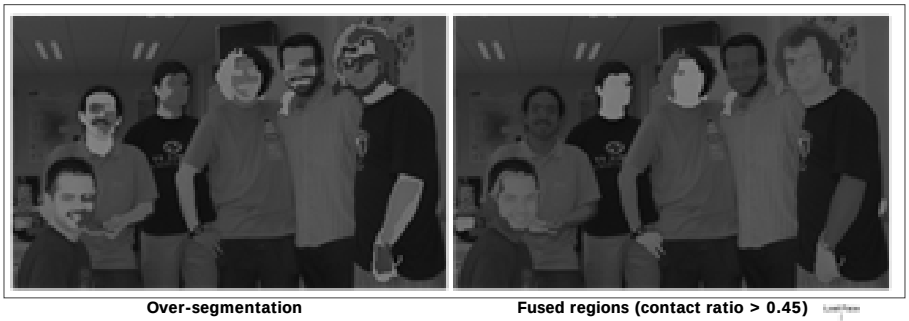
SAI (XIII): ROI filtering

- With the presented set of ROI descriptors it's possible to perform filtering of the first set of ROIs by imposing filtering constraint based on the characteristics of the type of ROIs we want to preserve avoiding the need of more complex segmentation and analysis of the image.
- In our case the ROI filtering is driven by the constraint of preserving human faces in an image.
- The ROI characteristics used for **model based filtering** are:
  - high density,
  - small distance between mass and geometric center
  - bounding box aspect ratio < 1



SAI (XIV): ROI filtering - preproccesing

In the case of oversegmentation (generic segmentation) there is the need of a previous ROI fusion



Over-segmentation

Fused regions (contact ratio > 0.45)

SAI (XV): ROI filtering - preprocessing

A first step of noisy ROIs' filtering is performed

- Binary segmentation: applying a size and density filtering to the set of ROIs in order to avoid small noisy ROIs.

Original ROIs (144 ROIs) **Noisy ROIs filtering (10 ROIs)**

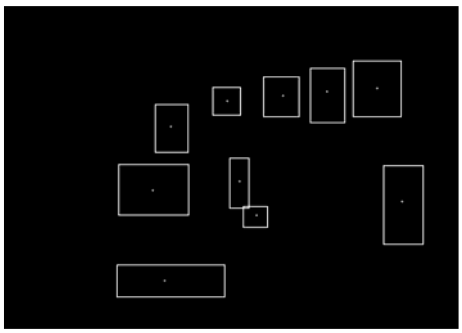
- Generic segmentation: filtering based on ROI descriptors

Descriptor	Constraint
Number of points	[1000, 3000]
Compactness	>0.6
Aspect Ratio	[0.333, 3]

Noisy ROIs filtering

SAI (XVI): ROI filtering – model filtering

In the next step the ROIs which doesn't fulfill the density and aspect ratio conditions are eliminated to avoid the 'non human face' regions.



Noisy ROIs filtering (10 ROIs)

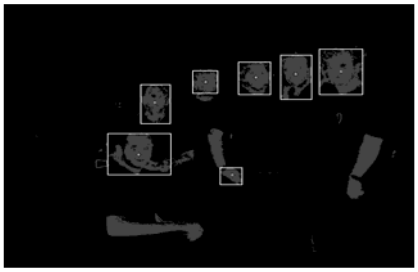
'Non human face' ROIs filtering

SAI (XVII): ROI Merging

After the filtering of ROIs, an optional merging process is applied in order to consider very close ROIs as unique ROIs.

The constraint imposed is a small distance between the mass center of the ROIs to ensure that the greater possible number of points of the ROIs are close.

The objective of ROIs merging is to connect fragments of a face (e.g., due to glasses, beard, …) or grouping people.

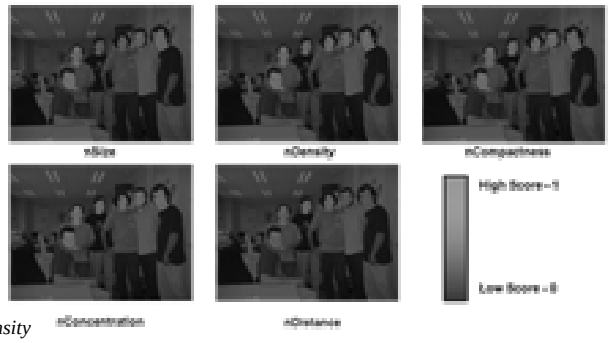


'Non human face' ROIs filtering

ROIs merging

SAI (XVIII): ROI Scoring (I)

The relevance value of each ROI is obtained by linear weighted combination of different features (descriptors)



$$rSize = nSize \cdot wSize$$

$$rDensity = nDensity \cdot wDensity$$

$$rCompactness = nCompactness \cdot wCompactness$$

$$rConcentration = nConcentration \cdot wConcentration$$

$$rDis\tan ce = nDis\tan ce \cdot wDis\tan ce$$

$$r = rSize + rDensity + rCompactness + rConcentration + rDis\tan ce$$

- Features and weights are application dependent

SAI (XIX): ROI Scoring

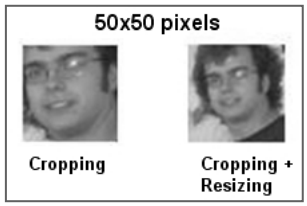
Measurement	Weight
Distance	0.32
Concentration	0.32
Size	0.18
Density	0.09
Compactness	0.09

SAI (XX): Image adaptation

Once the set of ROIs is scored it's possible to perform different image adaptations based on relevance

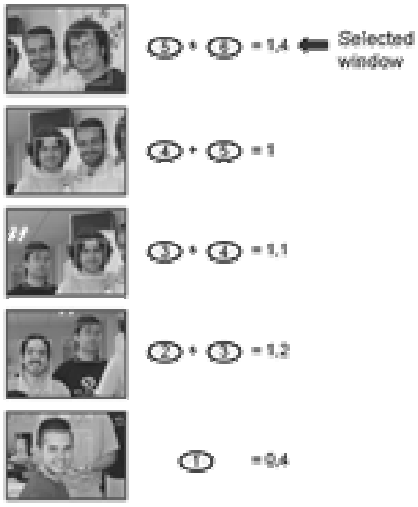
- Semantic Image cropping
- Semantic quality adaptation
- Privacy
- Other combinations of quality based adaptation
- …
- Image2Video transmoding

SAI (XXI): Image adaptation - Cropping



Cropping + centering

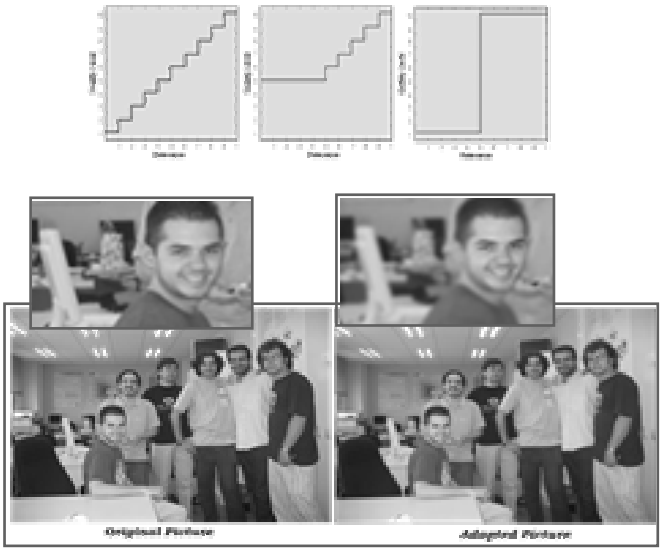
SAI (XXII): Image adaptation – Image Cropping



ROIs relevances

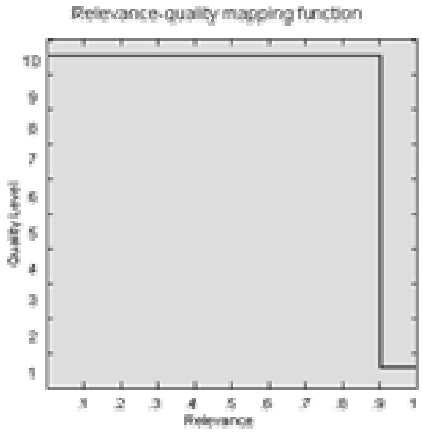
SAI (XXIII): Image adaptation – quality adaptation

SAI (XXIV): Image adaptation – quality adaptation



File size reduction of 30 %

SAI (XXV): Image adaptation – quality adaptation privacy



SAI (XXVI): Image adaptation – quality adaptation

SAI (XXVII): Conclusions

Semantic Adaptation of Images allow to lose less relevant information when adapting images to different terminals and networks.

The presented algorithms allow quick image adaptation by means of using rough segmentation techniques (segmentation is the more resource consuming step) complemented by ROI management (filtering, merging, scoring, …)

Several approaches to final adaptation based on relevance values of ROIs have been discussed.

SAI (XXVIII): References

★ Víctor Valdés, José M. Martínez, “Fast Image Adaptation driven by people presence”, Proceedings of Second European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, EWIMT 2005, Londres, Reino Unido, Noviembre-Diciembre 2005, pp. 199-204.

★ Víctor Valdés, “Adaptación de contenido basada en regiones de interés”, Trabajo de Iniciación a la Investigación, Programa de Doctorado en Ingeniería Informática y de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid. Septiembre 2006.

I2V (I): Introduction

Communication networks and Internet → Unique universal network

Different client devices, with different user preferences

Slow but steady convergence of processing power in desktop and mobile devices

Limitation: Screen size

Approaches to present large images on mobile displays:

- Downsampling
- Cropping
- *Recomposition*
- **Image2Video**

I2V (II): Existing adaptations of images to small displays

Downsampling

Blind and Semantic Cropping

Image2Video

I2V (III): Characteristics of Image2Video adaptation

Trading of space and time

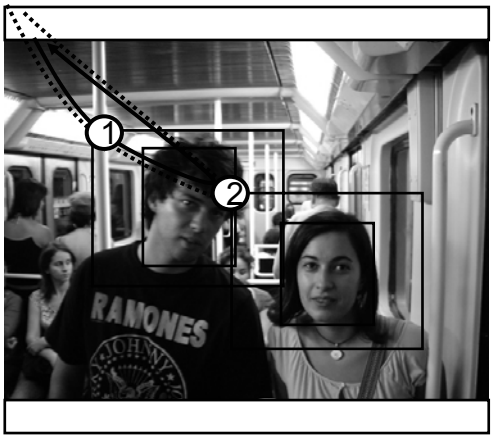
The viewer is only able to concentrate on a single attention object

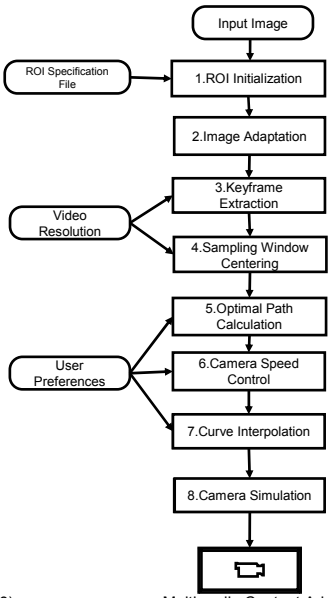
Maximization of the information fidelity (i.e. the information of the original image is maintained)

The information is concentrated in the attention objects/regions of interest (ROIs)

Usual attention objects: Faces, text and saliencies, although it depends strongly on each application.

I2V (IV): Image2Video flow





I2V (V): results

I2V (VI): Conclusions

Image2Video transmoding provides a new alternative for automatically browsing images in detail or in small displays

Depending on the objects relevant, different analysis techniques for ROI definition can be used

Also user preferences can be taken into account with respect to the montage of the video

1. Video resolution
2. Camera speed
3. Time stopped at each ROI
4. Maximal zoom
5. Curvature of the Catmull-Rom interpolation
6. ROI specification or image previewing
7. Automatic or manual sorting of the ROIs
8. Codec and bitrate

I2V (VII): References

★ Fernando H. Barreiro Megino, “Transformación de imagen a vídeo”, Proyecto Fin de Carrera, Univ. Autónoma de Madrid, Escuela Politécnica Superior, Octubre 2007.

OLSVA (I): Introduction

OL … On-line

- Analysis and Selection Techniques
  - Off-Line: Techniques applied to the whole available video.
  - **On-Line**: Frame-by-frame / GoP-by-GoP / shot-by-shot operation.

S … Semantic

- **Semantic adaptation** allows to preserve more relevant information during summarization
  - Multimedia analysis for content understanding

VA … Video Abstraction

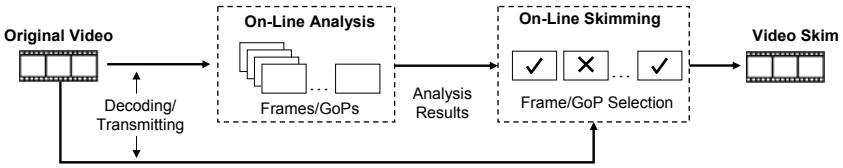
- **Video summarisation** is usually used for meaning **video abstraction** that is mainly divided in:
  - keyframe based abstraction: ***video summarization***

    video units –e.g., shots, GoPs- based abstraction: ***video skimming***

    Nevertheless, the analysis may work frame-by-frame, whilst continuity of frames can be accomplished in a post-processing stage prior to video generation/presentation.
  - Here we will use video summarization generically sometimes ;)

OLSVA (II): On-line summarization approach

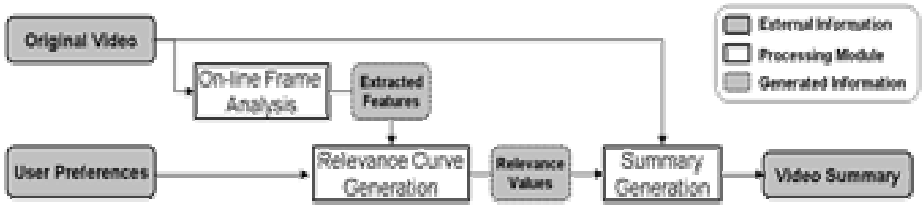
Analysis and frame selection is performed as the original video is being received (or decoded)



- Advantages:
  - Provides adaptation mechanisms for improving the efficiency of the video skimming task.
  - Enables the possibility of generating videos skims of broadcasted content as it is being received or watching video skims from stored video content as they are being generated.
- Limitations:
  - Given a specific time instant there is not information available about the incoming video to take into account in the analysis process.
  - If the size of the original video is unknown it can be impossible to produce a specific-length video skim.
  - Once a video unit has been written in the output it is not possible to eliminate it.

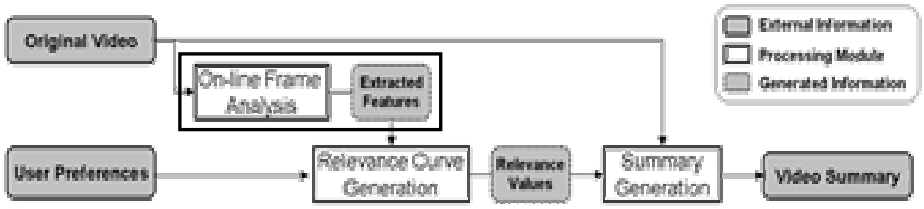
OLSVA (III): Proposed On-line Summarization Framework

On-Line Summarization Framework based on relevance curves:

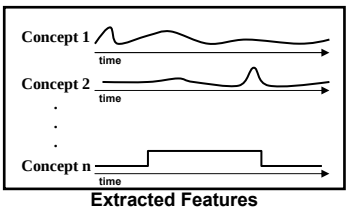


- Provides support for user preferences consideration.
- Relies on on-line analysis techniques.
- The model supports the generation of keyframes or video-skim modalities.

OLSVA (IV): On-Line Analysis

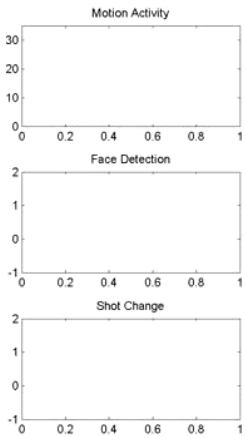


- On-line at the analysis phase is heavily related with "real-time" analysis
  - Simple features
  - Efficient algorithms
  - Analysis in the compressed domain
    - Compressed domain data (e.g., DCT coeff, MV)
    - Partial decoding
- Must output one or more semantic curves:
  - People or Face Presence
  - Motion Activity
  - Content Classification
  - Audio Energy
  - …



Extracted Features

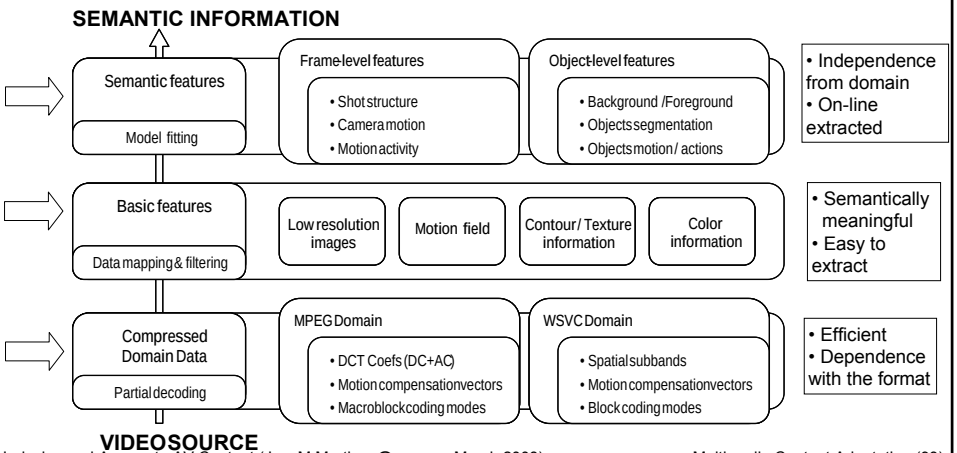
OLSVA (V): On-line features extraction



OLSVA (VI): On-line analysis framework

Analysis can be seen as a hierarchical approach

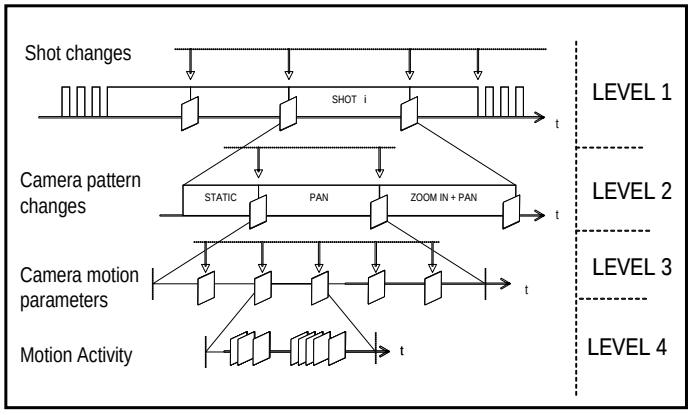
- Allows independence of underlying coding format (just one layer is coding format dependent)



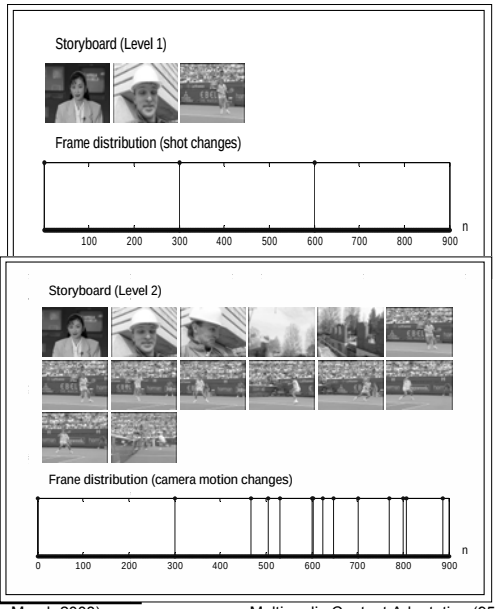
OLSVA (VII): Selection of frames based on generic semantic features

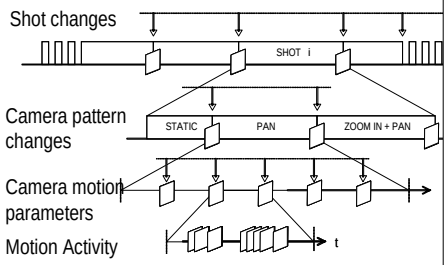
Here we present a model that directly selects frames without generating a (continuous) relevance curve

- Can be expressed as a quantified relevance curve (selection curve)

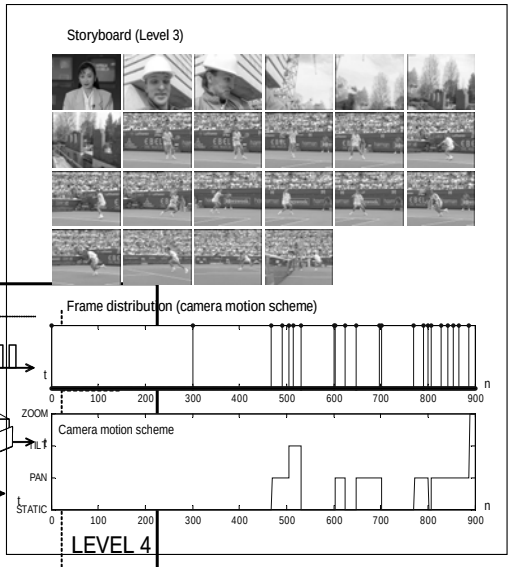


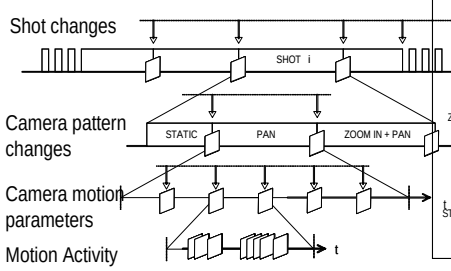
OLSVA (VIII): Selection of frames based on L1&2 features



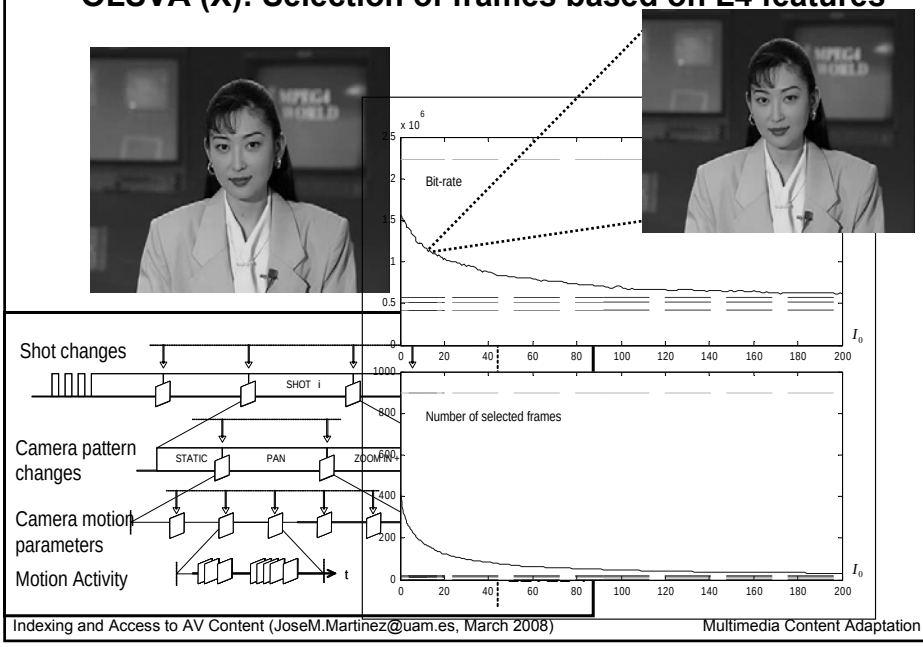


OLSVA (IX): Selection of frames based on L3 features

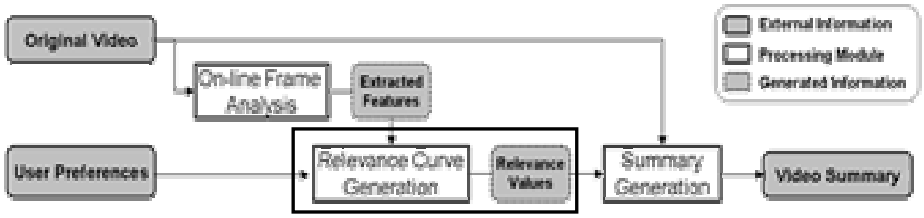




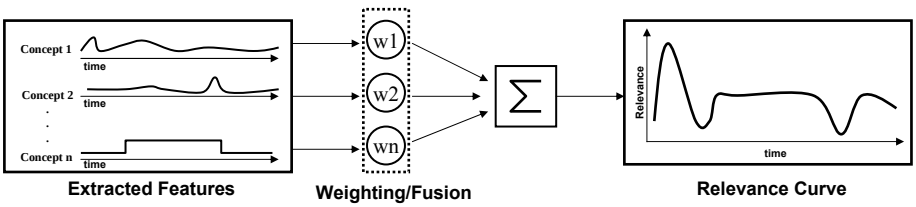
OLSVA (X): Selection of frames based on L4 features



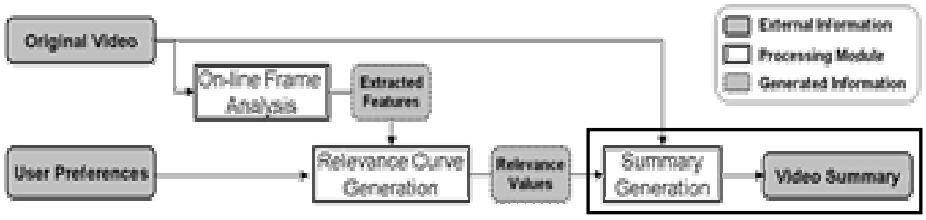
OLSVA (XI): Relevance curve generation



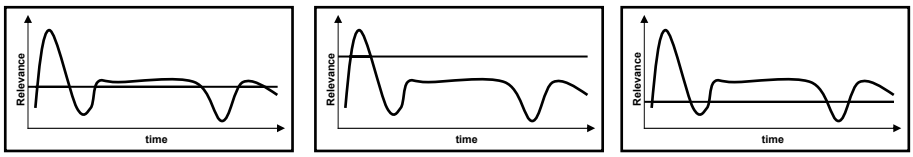
- The relevance curve: represents the relevance (for the user or application domain) of each video portion.
- Analysis obtained concepts are weighted (according to the user preferences or application domain) to obtain the final relevance curve which represents the relevance of each video unit (e.g., frames, shots, GoPs)



OLSVA (XII): Summary generation



- Performs frame selection to generate the final video summary.
- The target is to maximize the final relevance value associated to the selected video fragments composing the summary.
- Summary can be generated by setting a relevance threshold for the frame selection
  - This simple basic approach produces several annoying effects in video skims (mainly discontinuities and spurious effects) that can be solved by **on-line post-processing (PP) techniques**

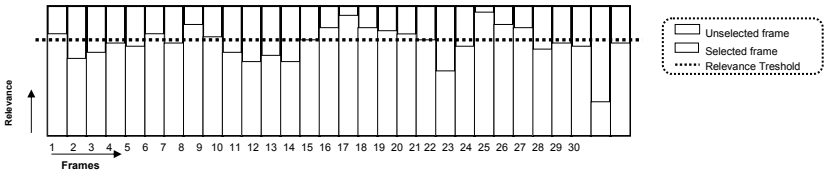


Fragment Selection (a) **Fragment Selection (b)** **Fragment Selection (c)**

OLSVA (XIII): Storyboard summary

OLSVA (XIV): "Simple threshold" Video Skimming

Thresholding Relevance Value



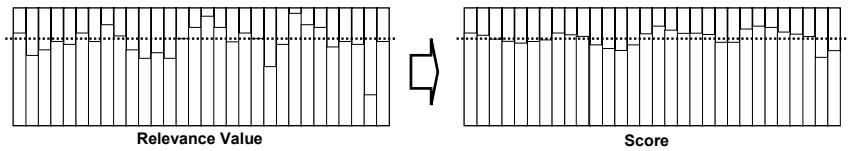
- Simple Solution
- Produces discontinuity effects on output video summary

Original Video **Face Detection Summary** **Motion Activity Summary**

OLSVA (XV): Relevance Smoothing PP

$$Score(i) = Score(i-1) \cdot SF + \mathrm{Re}levanceCurve(i) \cdot (1 - SF)$$

SF = Score Fade



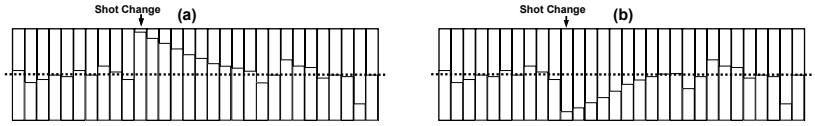
Relevance Value **Score**

Motion Activity Based Summary **Smoothed Score Summary**

OLSVA (XVI): Shot Boundary Segments Elimination PP

$$Score(i) = Score(i-1) \cdot SF + (\mathrm{Re}\,levanceCurve(i) + SCB) \cdot (1 - SF)$$

SF = Score Fade; SCB = Shot Change Bonus

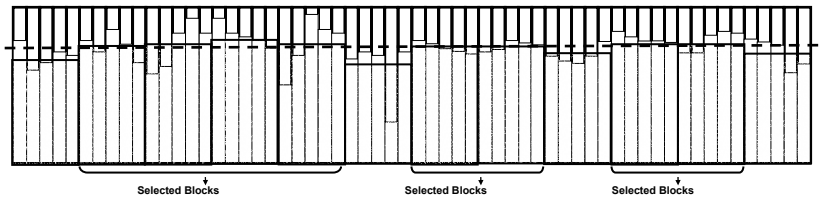


Score Value

Smoothed Score Summary

SCB+Smoothed Score Summary

OLSVA (XVII): Fixing Minimum Video Segment Size PP

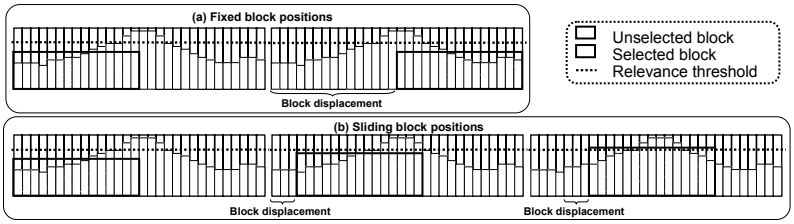


•**ALLOWS IMPROVEMENT ON THE VIDEO SUMMARY CONTINUITY**

SCB+Smoothed Summary

Minimum Block 35 Frames Summary

OLSVA (XVIII): Optimizing BoF Positioning PP



•**ALLOWS IMPROVEMENT ON THE VIDEO SUMMARY COVERAGE**

Minimum Block 35 Frames Motion Summary

5-Frame Slide Block Summary

OLSVA (XIX): Rate Control Mechanism

Working in on-line mode makes impossible to target to a summarization fixed length. (There is not a priori information about the original video length).

But it is possible to target to a summarization length defined as a portion of the received video (*resultingTimeRatio*)

Nevertheless the on-line operation mode implies the lack of information about the incoming video which makes to achieve a specific *resultingTimeRatio* length rate a difficult task.

$$resultingTimeRatio = \frac{length(summary)}{length(original)}$$

OLSVA (XX): Fixed Relevance Threshold

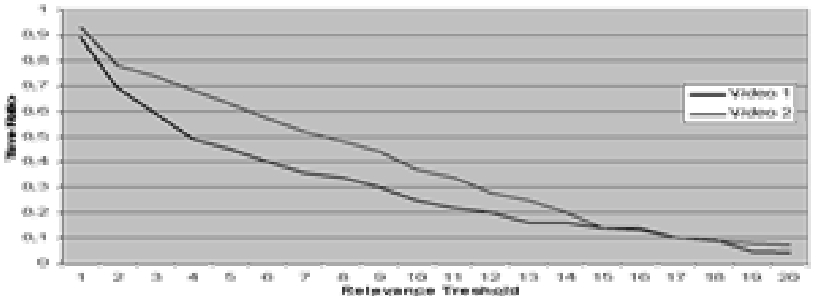
A fixed relevance threshold is set.

Previous described techniques for summary inclusion and smoothness are applied.

Two original news bulletin videos (10000 frames) are used, using the motion activity as relevance value.

Differences between the output summaries will depend on the specific particularities of each video, kind of content, etc.

There is not guarantee of obtaining a specific *resultingTimeRatio*.



OLSVA (XXI): Adaptive Relevance Threshold

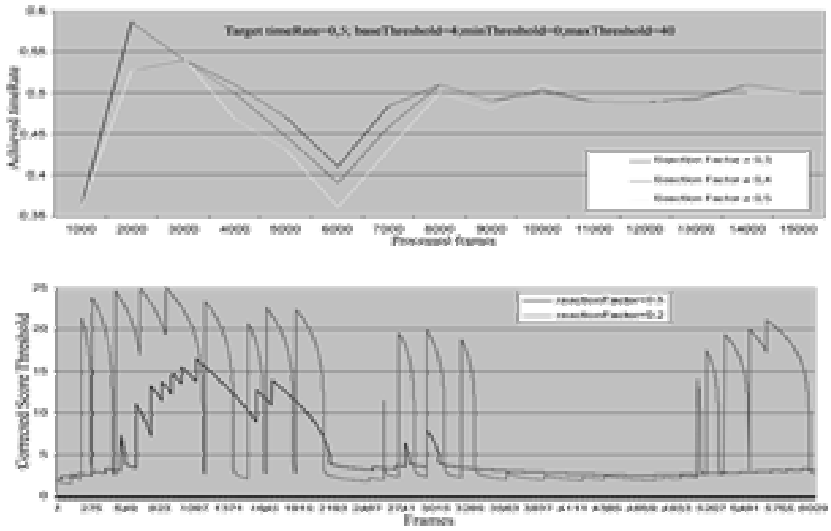
It is necessary to gain control over the output rate of the summarization *resultingTimeRatio.*

The relevance threshold will be dynamically recalculated while the summarized video is being generated.

As there is not available information about the incoming video the correction in the relevance threshold must be calculated using the available summary information in the following steps:

1. Current summary size: $$CurrentTimeRatio = \frac{writtenFrames}{processedFrames}$$
2. Correction Factor calculation:
$$CF = \left[\text{abs}(targetTimeRatio - CurrentTimeRatio)\right]^{reactionFactor}$$
3. If *CurrentTimeRatio* > objective *TimeRatio*
$$resultThreshold = BaseThreshold + (maxThreshold - BaseThreshold) \cdot CF$$
4. If *CurrentTimeRatio* < objective *TimeRatio*
$$resultThreshold = BaseThreshold - (BaseThreshold - minThreshold) \cdot CF$$

OLSVA (XXII): Adaptive Relevance Threshold - Reaction Factor Effects-



OLSVA (XXIII): Adaptive Relevance Threshold - Reaction Factor Effects-

Reaction Factor

Target Rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.2	7.48	8.49	9.15	9.24	9.72	9.98	9.80	9.6
0.4	5.74	6.79	7.36	7.85	7.85	7.99	7.89	7.99
0.6	6.13	6.59	6.68	6.78	6.74	6.74	6.74	6.80
0.8	5.85	6.05	6.2	6.4	6.49	6.5	6.58	6.58

Mean Summary Relevance

Reaction Factor

Target Rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.2	0.19	0.21	0.23	0.23	0.25	0.25	0.27	0.3
0.4	0.37	0.36	0.37	0.39	0.40	0.41	0.42	0.41
0.6	0.55	0.54	0.54	0.55	0.56	0.56	0.56	0.56
0.8	0.66	0.64	0.63	0.62	0.61	0.61	0.60	0.60

Resulting Summary Size

*More Abrupt Relevance Threshold Changes
*Better summary size results
*Worse frame selection

*Smoother Relevance Threshold Changes
*Worse summary size results
*Better frame selection

LOW *ReactionFactor* Value

HIGH *ReactionFactor* Value

In order to get the best results, the different configuration variables (*baseThreshold, minThreshold, maxThreshold* and *reactionFactor*) must be set depending on the nature of the original video and a statistical study of the Relevance value calculation

OLSVA (XXIV): Conclusions

A generic system for on-line video summarization based on relevance curves has been proposed.

The presented algorithms can deal with any kind of extracted features while providing support for user preferences inclusion.

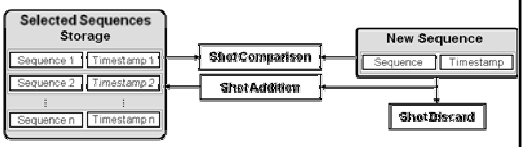
The system includes mechanism for on-line frame selection aimed to maximize de output summary relevance while trying to maintain smoothness in the output video summary.

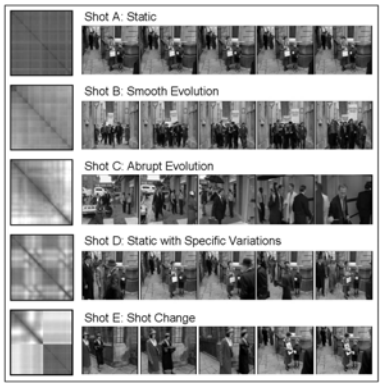
An adaptive rate control mechanism has been proposed enabling the possibility of controlling the size of the output summary even with limitations of the on-line operating mode.

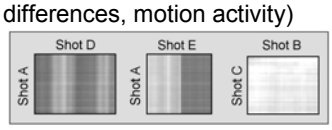
OLSVA (XXV): Current and future Work – redundancy filtering

Another approach to video summarization is redundancy filtering

- On-line redundancy filtering stores information about the already selected video units and use them in order to filter similar (near) video units, therefore enhancing semantic coverage
- Relevance curves are substituted by some (timed) similarity measure among video units (e.g., Histogram differences, colour layout differences, motion activity)







OLSVA (XXVI): Current and future work – Composite Video Summaries Presentation

Additionally to traditional keyframe presentations (storyboards, fast forward) there are innovate proposals like video posters.

With respect to video skims presentations less innovation has been proposed (besides "multiple simultaneous views").

- With some little additional delay (depending of the higher semantic unit –e.g., scene, news item-) on-line summaries may be presented in a composite layout allowing further visual semantic coverage (including also graphics) as well as audio synchronization.
- Further multimedia analysis is requiered

OLSVA (XXVII): References

- ★ Jesús Bescós, José M. Martínez, Luis Herranz, Fabricio Tiburzi, "Content-driven Adaptation of On-line Video", Signal Processing: Image Communication, 22(7-8):651-668, Agosto-Septiembre 2007.
- ★ Víctor Valdés, José M. Martínez, "On-line Video Skimming Based on Histogram Similarity", en Proceeding of the ACM Multimedia 2007 Workshop on TRECVID Video Summarization, Augsburg, Alemania, Septiembre 2007.
- ★ Víctor Valdés, José M. Martínez, "Post-processing techniques for on-line adaptive video summarization based on relevance curves", en Semantic Media and Digital Media Technologies-SAMT07, B. Falcidieno, M. Spagnuolo, Y. Avrithis, I. Kompatsiaris, P. Buitelaar (eds.), Lecture Notes in Computer Science, Vol. 4816, Springer Verlag, 2007, pp.144-157.

MCA: some experiences

SASV (I): Introduction

S … Semantic

- Usually in adaptation we lose some information
  - Try to preserve those parts more "semantically" relevant
- Requires analysis of the content
  - Previously stored (metadata) or on-line processed (usually features from the compressed domain)

A … Adaptation

- Not only abstraction (temporal dimension), but also resolution and quality adaptation

SV … Scalable Video

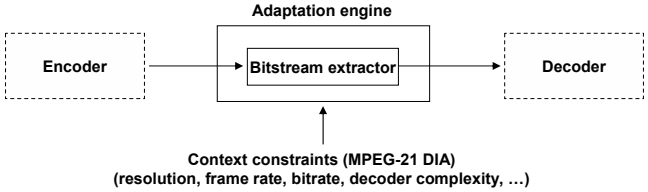
- Scalable formats allow to perform adaptation without the need of decoding/reencoding
  - Besides the required decoding for the analysis phase

    Once if stored as metadata (e.g. MPEG-21 (g)BSD)

SASV (II): Adaptation Engine for SV

Adaptation engine for scalable video

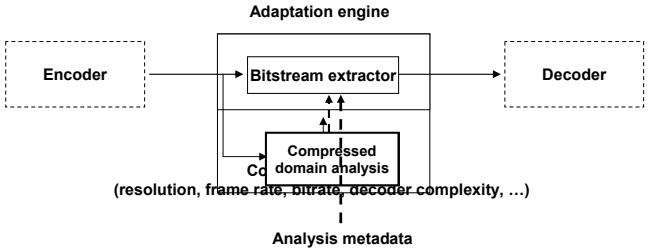
- Context aware (MPEG-21 Usage environment descriptions)
- Content-blind (independent of the content)
- Advantage: adaptation=bitstream extraction (**very fast)**
  - ¿Can we take advantage of the semantics in the content?
  - ¿Can we keep the efficiency?



SASV (III): Adaptation Engine for SV

Semantic adaptation engine for scalable video

- Context aware (MPEG-21 Usage environment descriptions)
- Content aware
  - Bitstream extraction guided by content (analysis)
- Efficiency relies on efficient analysis
  - Analysis stored previously as metadata
  - Compressed domain analysis (shot change detection, activity)



SASV (IV): Analysis

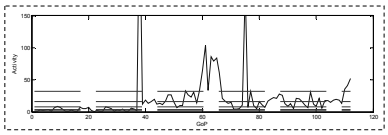
Analysis in the framework

- Efficient analysis
  - Compressed domain: depends on the codec
- But for most codecs there are several common features that can be extracted very fast
  - Motion vectors
  - Low frame rate and low resolution versions

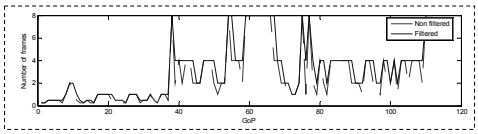
SASV (V): Activity based skimming

Skimming curve

- Indicates the temporal level to be selected in each GoF
- M possible temporal levels
  - Assuming T temporal decompositions: T+1 temporal levels in the scalable bitstream ({1,2,4,8, … 2^T} frames in each level)
  - If M>T+1 lower temporal level is achieved skipping GoFs
- Obtained by thresholding an activity curve



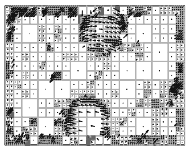
GoF activity



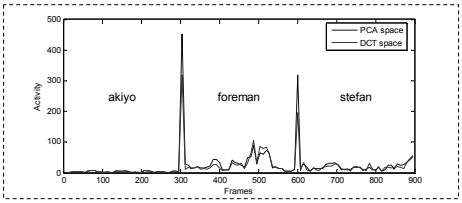
Skimming curve

SASV (VI): Activity based skimming

Experiments on GoF activity



- Using motion vectors
  - MPEG-7 intensity of motion activity
    - Problems: highly dependent on the codec implementation
    - Sometimes is still computationally expensive
- Using low resolution versions
  - Lowest frame rate, lowest frame size : 1 frame per GoF
  - PCA analysis was proposed as an efficient measure of activity (Li et al 2005)
    - Problem: compute the basis depends on the whole sequence (needs all frames)
  - Using 1-D DCT
    - Results are very similar to PCA analysis
    - Adavantage: the basis is always the same (DCT), suitable for online activity analysis



SASV (VII): Scalable video based summarization

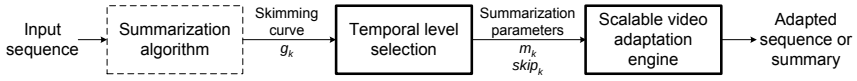
Video abstraction/summarization as a special case of video adaptation (structural adaptation)

As an application of the previous framework

- Dynamic selection of the temporal level
  - o Skimming curve: indicates the temporal level to be selected in each GoF
  - o Summarization parameters

    $skip_k$ (*true* if the GOF is skipped)

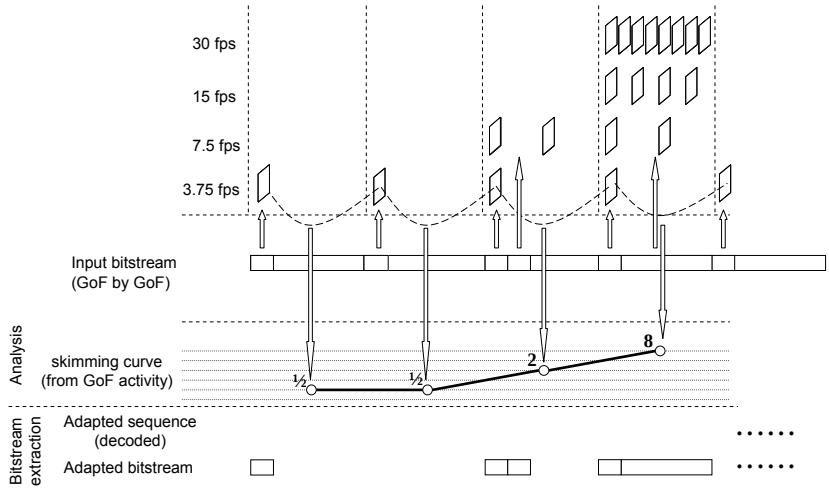
    m_k (temporal level of the GOF)



SASV (VIII): Scalable video based summarization

Summarization using scalable video

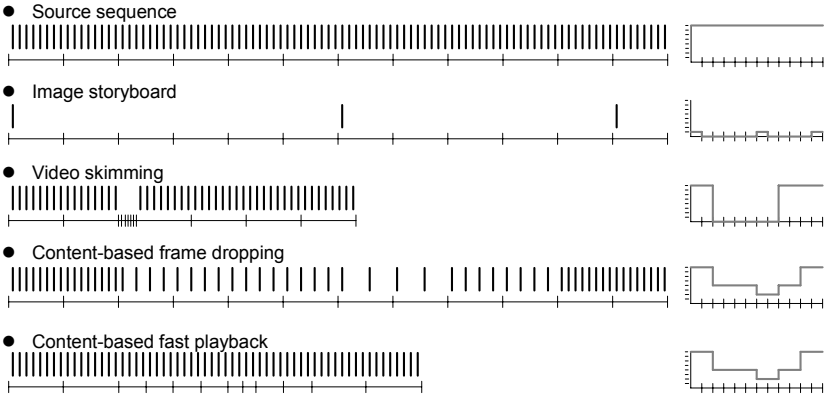
- Example: online activity analysis



SASV (IX): Scalable video based summarization

Skimming curve examples

- Source sequence
- Image storyboard
- Video skimming
- Content-based frame dropping
- Content-based fast playback



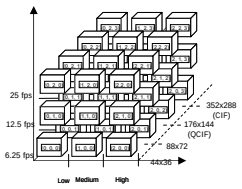
SASV (X): Including the usage environment

At this point, there is no adaptation to usage environment (unconstrained)

Fully scalable video adaptation using MPEG-21 DIA

- Optimization problem: Determine the adaptation coordinates per GoF that maximize a GoF quality measure given the usage environment constraints
  - Independent variables: (NTEMP, NSPATIAL, NQUAL)
  - Measure: PSNRGOF
  - Constraints: (display_width, display_height, average_bitrate…)
- Problem

  maximize PSNRGOF=f(NTEMP, NSPATIAL, NQUAL; NGOF)
  subject to
  FRAMEWIDTH ≤ *display_width*
  FRAMEHEIGHT ≤ *display_height*
  FRAMERATE ≤ *display_refresh_rate*
  BITRATE_BUDGET ≤ *average_bitrate*



SASV (XI): Including the usage environment

Including semantic constraints

- Include results from analysis in the optimization problem
  - A new independent variable: INCLUDE_GOF
  - New semantic constraints: *analysis_temporal _level*, *skip*

    Obtained from the skimming curve
- Problem

maximize PSNRGOF=f(NTEMP, NSPATIAL, NQUAL, ***INCLUDE_GOF****; NGOF)*
subject to
FRAMEWIDTH ≤ *display_width*
FRAMEHEIGHT ≤ *display_height*
FRAMERATE ≤ *display_refresh_rate*
BITRATE_BUDGET ≤ *average_bitrate*
NTEMP ≤ *analysis_temporal_level*
INCLUDEGOF = NOT(*skip*)

SASV (XII): Results

Source sequence
(352x288 30 fps)

Adapted sequence
(176x144 fast browsing at 30 fps)

Adapted sequence
(176x144 dynamic frame rate decoder)

SASV (XIII): Conclusions

Framework combining scalability and semantic analysis for adaptation
- Compressed domain for efficient analysis
- Bitstream extraction for efficient adaptation
- Generic framework that could be used with other fully scalable video codecs

Application to different summarization approaches
- Experiments on video skimming, fast playback and image storyboard

Integrated into the standard MPEG-21 DIA framework

SASV (XIV): References

★ Luis Herranz, “Integrating semantic analysis and scalable video coding for efficient content-based adaptation”, Multimedia Systems, 13(2):103-118, Agosto 2007.

★ Luis Herranz, José M. Martínez, “Use Cases of Scalable Video Based Summarization and Adaptation within MPEG-21 DIA”, en Semantic Media and Digital Media Technologies-SAMT07, B. Falcidieno, M. Spagnuolo, Y. Avrithis, I. Kompatsiaris, P. Buitelaar (eds.), Lecture Notes in Computer Science, Vol. 4816, Springer Verlag, 2007, pp.256-259.

Index

Multimedia Content Adaptation: a short overview
- Introduction
- Content adaptation dimensions
- Related Standards
- Conclusions
- References

Multimedia Content Adaptation: some experiences
- The CAIN adaptation engine
- Semantic adaptation of images
- Image2Video transmoding
- On-line semantic video abstraction
- Semantic adaptation of scalable video
- …

Annex: MPEG-7 and MPEG-21 metadata for content adaptation

Grupo de Tratamiento de Imágenes

Index

Multimedia Content Adaptation: a short overview
- Introduction
- Content adaptation dimensions
- Related Standards
- Conclusions
- References

Multimedia Content Adaptation: some experiences
- The CAIN adaptation engine
- Semantic adaptation of images
- Image2Video transmoding
- On-line semantic video abstraction
- Semantic adaptation of scalable video
- …

Annex: MPEG-7 and MPEG-21 metadata for content adaptation

MPEG-7 and MPEG-21 metadata for content adaptation

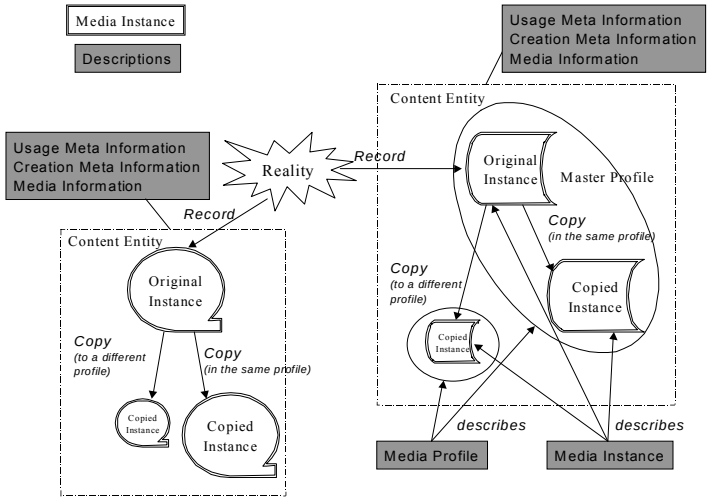
MPEG-7

- Content Description
  - Description of multimedia formats, archival information, …
  - Description of variations and summaries
  - Description of transcoding hints
  - Description of semantic content

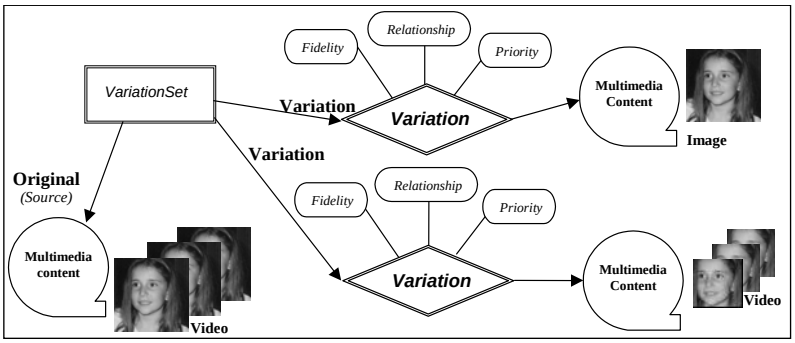
MPEG-21: Part 7 – Digital Item Adaptation

- Description of terminal, network, user and environmental conditions
- Tools for multimedia resources adaptation
  - XML representation of the structure of scalable formats for format independent adaptation
  - Relation ships among usage environment constraints, feasible adaptations and quality of results (at QoS level)
  - Tools for metadata adaptation

MPEG-7: *Media Profiles*



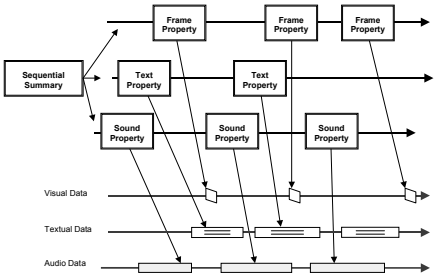
MPEG-7: Variations

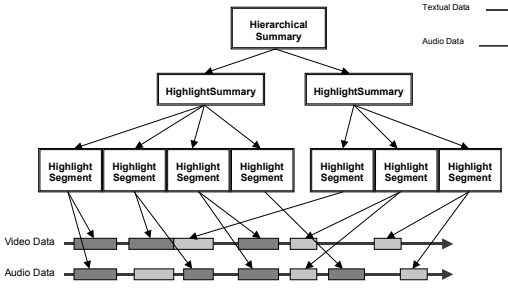


MPEG-7: Summaries

Sequential

Hierarchical





MPEG-7: *Transcoding Hints*

MediaTranscodingHintsType	Transcoding hints of the Media Profile.
MotionHint	Motion hints for a transcoder.
MotionRange	Motion range for a transcoder.
motionRangeXLeft	Indicates the recommended search range for horizontal motion vectors to the left.
motionRangeXRight	Indicates the recommended search range for horizontal motion vectors to the right.
motionRangeYDown	Indicates the recommended search range for vertical motion vectors to the bottom.
motionRangeYUp	Indicates the recommended search range for vertical motion vectors to the top.
motionUncompensability	Describes the amount of new content in the corresponding segment (this descriptor applies to descriptions attached to video segments). The motionUncompensability takes values from 0.0 to 1.0, where 0.0 indicates no new content and 1.0 indicates significant change in content.
motionIntensity	Describes the motion intensity in a segment. The motionIntensity takes values from 0.0 to 1.0, where 0.0 indicates low motion intensity and 1.0 indicates significant motion intensity.
ShapeHint	Shape hints for the transcoder.
shapeChange	Describes the amount of shape change in the corresponding segment (this descriptor applies to descriptions attached to video segments. The ShapeHint takes values from 0.0 to 1.0, where 0.0 indicates that no change has occurred and 1.0 indicates that all the pixels that define an object have been displaced.
avgNumNonTranspBlocks	Describes the average number of 16x16 blocks per frame containing at least one pixel with a non-zero alpha-map value.
CodingHints	Coding hints for the transcoder.
avgQuantScale	Describes the average quantization scale used to compress the media.
intraFrameDistance	Describes the distance between Intra-coded Frames, also known as N. A value of N=0 represents the case when N is infinite, for example, when the GOP has no I-frame (PBBPBBP...) or when there is only one I-frame at the start (IPPP...)
anchorFrameDistance	Describes the distance between anchor frames, also known as M, where an Anchor frame is defined as a frame that predictions are made from, for example, an I or P frame. A value of M=0 is forbidden.
difficulty	Describes the transcoding difficulty of the media. The difficultyHint takes values from 0.0 to 1.0, where 0.0 indicates the lowest difficulty and 1.0 indicates the highest difficulty.
importance	Describes the importance of the media. The importance takes values from 0.0 to 1.0, where 0.0 indicates the lowest importance and 1.0 indicates the highest importance.
spatialResolutionHint	Describes the maximum allowable spatial resolution reduction factor for perceptibility. The SpatialResolutionHint takes values from 0.0 to 1.0, where 0.5 indicates that the resolution can be reduced by half and 1.0 indicates the resolution cannot be reduced.

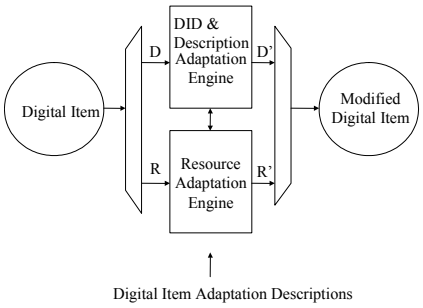
MPEG-21 Digital Item Adaptation

ISO/IEC 21000-7 Digital Item Adaptation (DIS) specifies the syntax and semantics of tools that may be used to assist the adaptation of Digital Items

- the Digital Item Declaration
- Resources and Descriptions (metadata) referenced by the declaration.

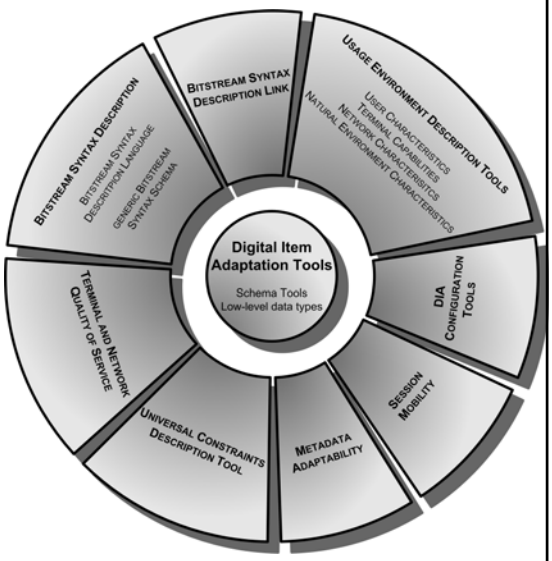
The tools could be used to satisfy transmission, storage and consumption constraints, as well as Quality of Service management by the various Users.

It is important to emphasize that the adaptation engines themselves are non-normative tools of DIA.



Overview of MPEG-21 DIA tools

- Schema tools and low level datatypes
- Usage Environment
- Resource adaptability (BSD)
- Terminal and Network QoS
- Universal Constraints
- Metadata Adaptability
- Session Mobility
- DIA configuration



DIA tools: Schema Tools and Low level datatypes

The schema tools provide uniform root elements for all DIA descriptions.

Low-level and basic datatypes can be used by several DIA tools independently.

DIA tools: Usage Environment

These tools provide descriptive information about the various properties of the usage environment, which originate from Users, to accommodate, for example, the adaptation of Digital Items for transmission, storage and consumption. They include:

- User Characteristics
- Terminal Capabilities
- Network Characteristics
- Natural Environment Characteristics

DIA tools: Usage Environment

User Characteristics description tools specify the characteristics of a User, including preferences to particular media resources, preferences regarding the presentation of media resources, and the mobility characteristics of a User. Additionally, description tools to support the accessibility of Digital Items to various users, including those with audio-visual impairments, are considered.

These tools are grouped in:

- User info, usage preferences and usage history
  - (imported from MPEG-7 MDS)
- Presentation preferences
  - Audio power, equalizer settings
  - Brightness, saturation, contrast
- Accessibility characteristics
  - Hearing thresholds at various frequencies in both ears
  - Type and degree of colour vision deficiencies
- Location characteristics
  - Mobility: angular changes and degree of random movement
  - Destination
  - To provide adaptative location-aware services

DIA tools: Usage Environment

Terminal Capabilities description tools specify the capability of terminals, including media resource encoding and decoding capability, hardware, software and system-related specifications, as well as communication protocols that are supported by the terminal.

These tools are grouped in

- Coded capabilities
  - Formats that can be decoded (profile@level)
  - Symmetric with MPEG-7 Media Format description tools
- Input-output capabilities
  - Display and audio output
  - Input devices
- Device properties
  - Power-related attributes
  - Storage
  - Data I/O characteristics

DIA tools: Usage Environment

Network Characteristics description tools specify the capabilities and conditions of a network, including bandwidth utilization, delay and error characteristics.

These tools are grouped in

- Network capabilities
  - Static attributes (e.g., maximum capacity, minimum guaranteed bandwidth)
- Network conditions
  - Dynamic parameters (available bandwidth, delay and error characteristics)

Natural Environment Characteristics specify the location and time of a User in a given environment, as well as audio-visual characteristics of the natural environment, which may include auditory noise levels and illumination properties.

These tools are grouped in

- Location and time
  - DI's location and time of usage
- Audiovisual environment
- Noise levels, noise frequency spectrum
  - Illumination characteristics

DIA tools: Resource adaptability (BSD)

Tools to assist with the adaptation of resources including the adaptation of binary resources in a generic way.

- BSD and gBSD

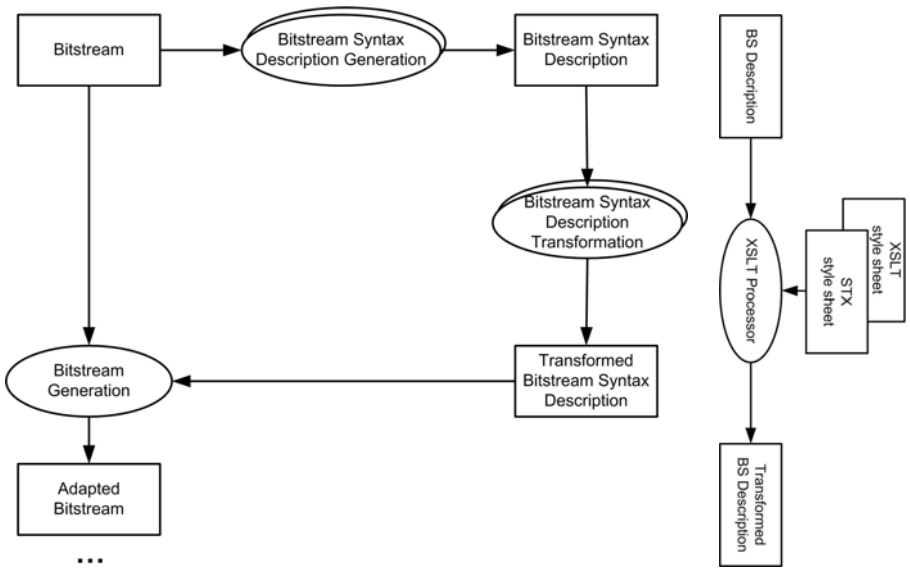
Additionally, tools that assist in making resource-complexity trade-offs and making associations between descriptions and resource characteristics for Quality of Service are targeted.

- BSDLink

DIA tools: Bitstreams Syntax Description

Bitstream Syntax Description tools comprise the third major category of Digital Item Adaptation tools. A BSD describes the syntax – in most cases, the high level structure – of a binary media resource. Using such a description, a Digital Item resource adaptation engine can transform the bitstream and the corresponding description using editing-style operations such as data truncation and simple modifications.

DIA tools: BSD "workflow"



DIA tools: BSD vs. gBSD

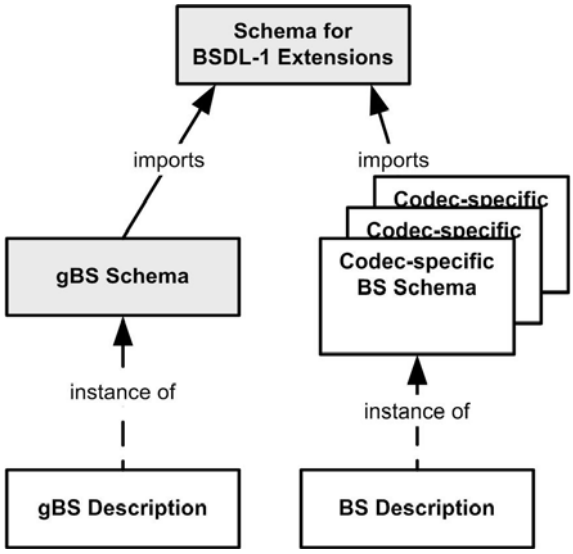
BSD

- Simple adaptation methods
  - Truncate or remove data
- Can be applied to a wide variety of (particular) coding formats

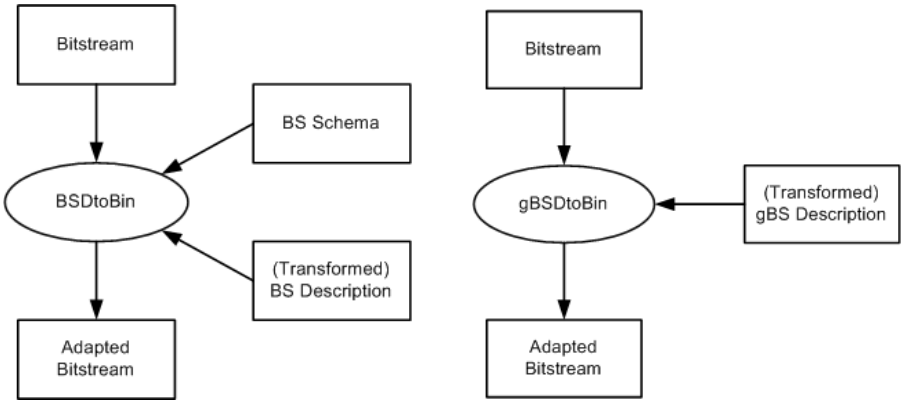
gBSD

- Describe any binary source in a codec-independent manner
- Provides means to associate semantic labels (markers) with the syntactical elements being described
- Allows to describe the bitstreams hierarchically
  - Groupings for efficient adaptation and addressing

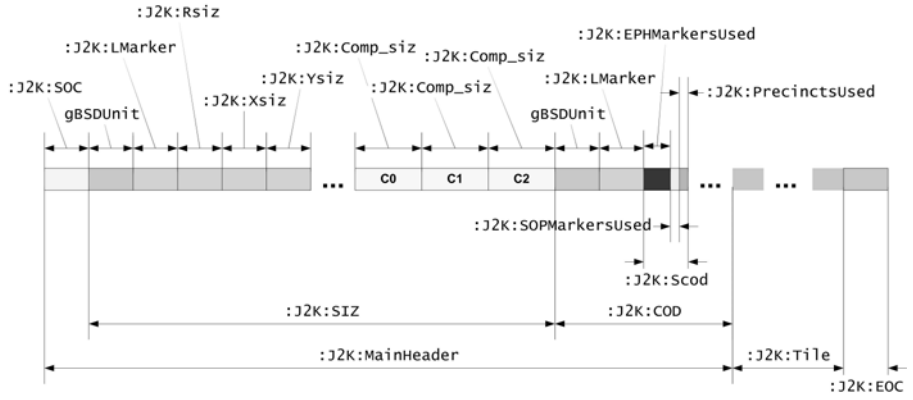
DIA tools: BSD vs. gBSD



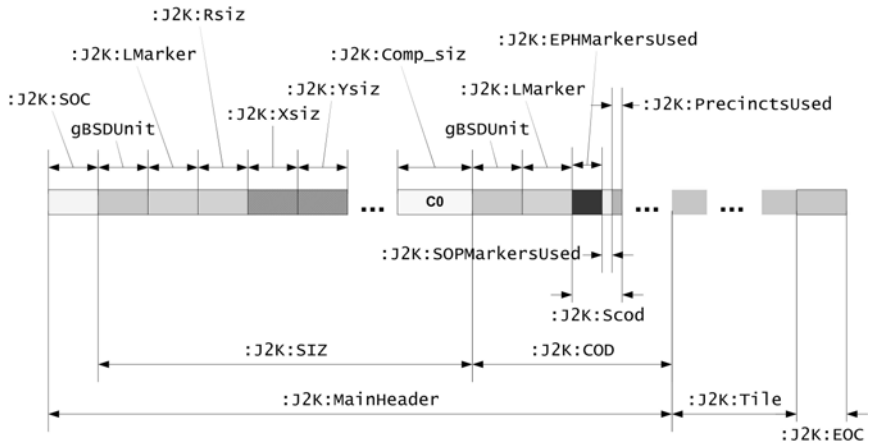
DIA tools: BSD vs. gBSD



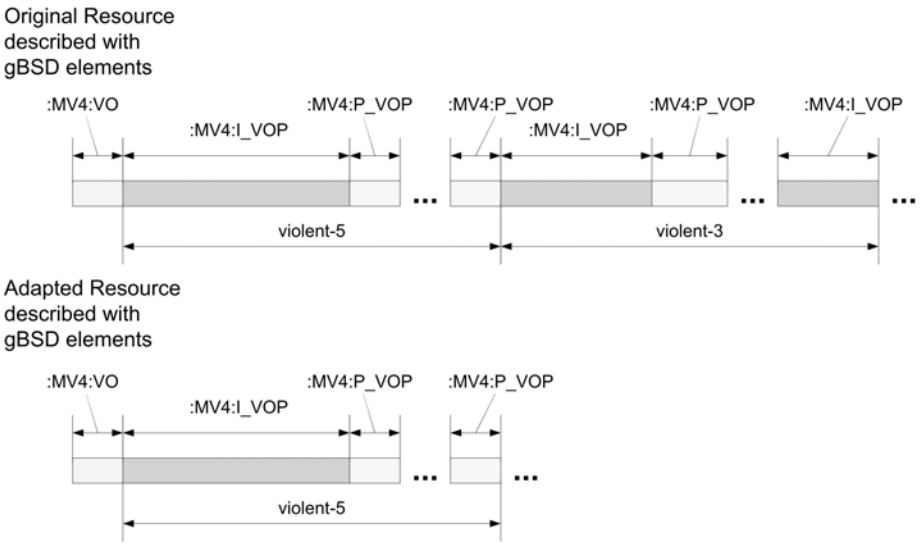
DIA tools: original resource with gBSD



DIA tools: adapted resource with gBSD



DIA tools: gBSD for semantic adaptation



DIA tools: BSDlink

BSDLink provides the facilities to create a rich variety of adaptation architectures based on tools specified within this part of ISO/IEC 21000, ISO/IEC 21000-2, and ISO/IEC 15398 among others.

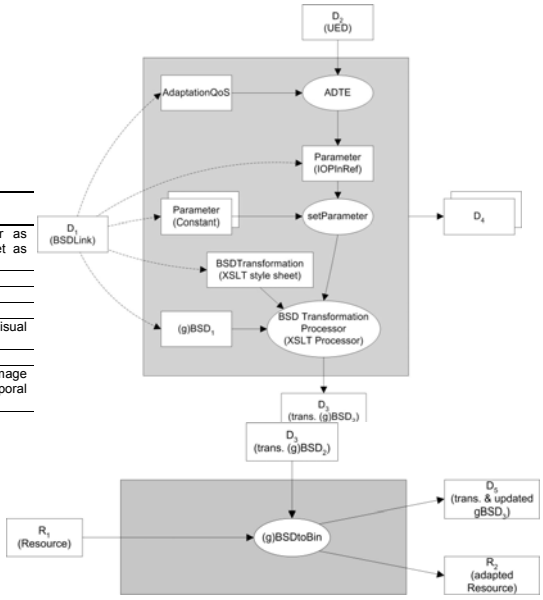
The BSDLink description tool includes references to:

- Steering adaptation description (AdaptationQoS)
- Bitstream and associated BSD
- BSD transformation
- Parameters and values of the steering adaptation description

DIA tools: BSDLink

Description Processing Engine

Short term	Example instantiation
D_1	BSDLink with an AdaptationQoS descriptor as steering descriptor and an XSLT style sheet as BSD Transformation.
D_2	Usage Environment Description (UED)
D_3	Transformed (g)BSD
D_4	Transformed BSDLink
R_1	Resource, e.g., JPEG2000 image, MPEG-4 Visual Elementary stream
D_5	Updated, transformed (g)BSD
R_2	Adapted Resource, e.g., smaller JPEG2000 image or grayscale JPEG2000 image, or temporal adapted MPEG-4 Visual Elementary stream.



Resource Processing Engine

DIA tools: Terminal and Network Quality of Service

The fourth category of tools is referred to as *Terminal and Network Quality of Service*. The tools specified in this category describe the relationship between QoS constraints (e.g., on network bandwidth or a terminal's computational capabilities), feasible adaptation operations satisfying these constraints and associated media resource qualities that result from adaptation.

This set of tools therefore provides the means to trade-off these parameters with respect to quality so that an adaptation strategy can be formulated and optimal adaptation decisions can be made in constrained environments.

DIA tools: Terminal and Network Quality of Service

```
<DIA>
  <Description xsi:type="AdaptationQoSType">
    <Header>
      <ClassificationAlias alias="AQoS" href="urn:mpeg:mpeg21:2003:01-DIA-AdaptationQoSCS-NS"/>
    </Header>
    <Module xsi:type="UtilityFunctionType">
      <Constraint iOPinRef="BANDWIDTH">
        <Values xsi:type="IntegerVectorType">
          <Vector>1510 1359 1200 1200 1071 1071 1071 941 814 814 814 1296 1000 1000 1000 842 744 909 712 600 396 359 331 293
255 217</Vector></Values>
      </Constraint>
      <AdaptationOperator iOPinRef="B_FRAMES">
        <Values xsi:type="IntegerVectorType"><Vector>0 0 0 1 0 1 2 0 0 1 2 1 0 1 2 1 1 2 2 2 2 2 2 2 2 2</Vector></Values>
      </AdaptationOperator>
      <AdaptationOperator iOPinRef="COEFF_DROPPING">
        <Values xsi:type="FloatVectorType">
          <Vector>0.0 0.1 0.21 0.09 0.3 0.2 0.0 0.4 0.5 0.44 0.31 0.0 0.35 0.27 0.08 0.4 0.5 0.2 0.4 0.5 0.0 0.1 0.2 0.3 0.4 0.5</Vector>
        </Values>
      </AdaptationOperator>
      <Utility iOPinRef="PSNR">
        <Values xsi:type="FloatVectorType">
          <Vector>34.47 33.56 32.48 31.40 31.58 30.82 28.62 30.27 29.10 28.57 27.53 31.94 30.69 30.15 28.33 29.04 28.11 28.01 27.03
26.49 23.44 23.36 23.29 23.18 23.02 22.87</Vector></Values>
      </Utility>
    </Module>
    <IOPin semantics=":AQoS:1.1.1" id="BANDWIDTH"/>
    <IOPin semantics=":AQoS:2.1" id="PSNR"/>
    <IOPin semantics=":AQoS:3.1.1" id="B_FRAMES"/>
    <IOPin semantics=":AQoS:3.1.2" id="P_FRAMES"/>
    <IOPin semantics=":AQoS:3.1.3" id="COEFF_DROPPING"/>
  </Description>
</DIA>
```


DIA tools: Universal Constraints and Metadata adaptability

The *Universal Constraints Description Tools* form the fifth category of tools which enables the possibility to describe limitation and optimisation constraints on adaptations.

The sixth category is referred to as *Metadata Adaptability*. This tool specifies hint information that can be used to reduce the complexity of adapting the metadata contained in a Digital Item. On the one hand, they are used for filtering and scaling, and on the other hand, for integrating XML instances.

DIA tools: Session Mobility and DIA Configuration

The seventh category of tools, *Session Mobility* specify how

- the configuration state information that pertains to the consumption of a Digital Item on one device is transferred to a second device. This enables the Digital Item to be consumed on the second device in an adapted way.
- to transfer the state of Digital Items from one User to another. Specifically, the capture, transfer and reconstruction of state information will be specified. This allows sharing DIs with the same (initial) choices and conditions.

Finally, the eighth category of tools is referred to as *DIA Configuration Tools*, which provides information required for the configuration of a Digital Item Adaptation Engine.

DIA tools: Amd. 1

In DIA Amd.1 new tools are being added:

- Conversion Capabilities (within Terminal Capabilites)
- Conversion Information
- Change Conditions
- Conversion Descriptors

Multimedia Content Adaptation

José M. Martínez

JoseM.Martinez@uam.es
Grupo de Tratamiento de Imágenes, Escuela Politécnica Superior
Universidad Autónoma de Madrid
http://www-gti.ii.uam.es

Index