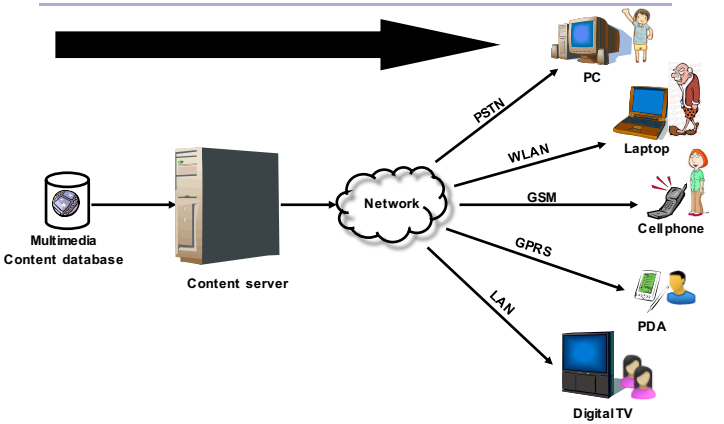
Multimedia Systems

Multimedia delivery

Dr. Yi-Zhe Song

Scenario



Today's agenda

- Motivation
- Adaptive content delivery: state of the art
  - Info pyramid
  - Scalable coding
  - Transcoding
- Video analysis for content adaptation
  - Semantic pre-filtering
  - Evaluation
- Image and video quality
  - Definition of the problem
  - Applications
  - Metrics

Motivation 1: personalisation

- Diffusion of network appliances
  - cellular phones
  - PDAs
  - hand-held computers, …
- Problem: how to adapt the media transmission to
  - Device capabilities
    - display
    - processing power
  - Network characteristics
    - bandwidth
  - User preferences
    - in accessing the same multimedia content

Motivation 2: enhanced applications

- Intelligent surveillance
  - Summarisation
  - Interpretation (event detection)
  - Indexing
  - Privacy preservation
  - Adaptation (mobile surveillance)
- Sport broadcasting
  - Instantaneous indexing
  - Avatars
  - Highlights detection

Queen Mary University of London

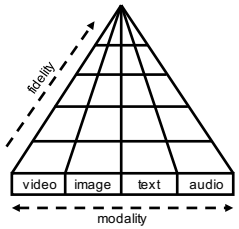
Today's agenda

- Motivation
- Adaptive content delivery: state of the art
  - Info pyramid
  - Scalable coding
  - Transcoding
- Video analysis for content adaptation
  - Semantic pre-filtering
  - Evaluation
- Image and video quality
  - Definition of the problem
  - Applications
  - Metrics

Queen Mary University of London

Info Pyramid

- Content is processed many times
  - Generation of variations
  - Server selects most appropriate variation
- Variations: different versions of media objects with
  - Different modalities
    - Video
    - Image
    - Text
    - Audio
  - Different fidelities
    - Summarised
    - Compressed
    - Scaled variations



Queen Mary University of London

Scalable coding

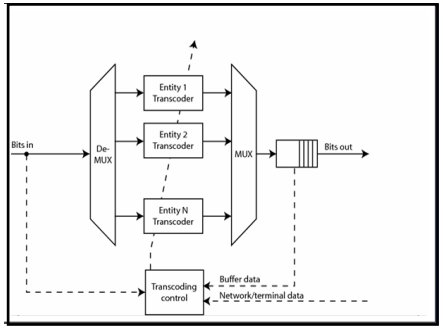
- Content is processed only once
  - Truncation of the bit-stream
- Scalability
  - Quality
    - Varying accuracies in the colour pattern (quantization)
  - Spatial
    - Varying spatial resolutions
  - Temporal
    - Varying temporal resolutions or frame rate
  - Frequency
    - Different frequency components in each layer (DCT or wavelets)

Queen Mary University of London

don't want to code everything

Transcoding

- Compressed content is converted into another signal with different properties



Queen Mary University of London

Transcoding

- Content-blind transcoding
  - Independent of the semantics in the content
  - Reduction of
    - Spatial resolution
    - Temporal resolution
    - Bit-rate
- Content-aware transcoding
  - Minimise the degradation of important image regions
    - bit-rate budget → controlled by *difficulty hint* of each object
    - frame skip → controlled by *shape hint*

Queen Mary University of London

Transcoding

- Intra-media transcoding
  - Do not change the media nature of the input signal
- Inter-media transcoding (trans-moding)
  - Process of converting one media input into another media format
    - Speech to text
    - Video to text
    - …

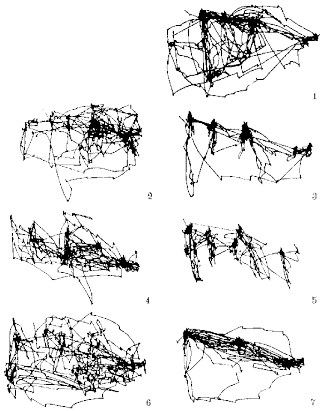
Queen Mary University of London

Today's agenda

- Motivation
- Adaptive content delivery: state of the art
  - Info pyramid
  - Scalable coding
  - Transcoding
- Video analysis for content adaptation
  - Semantic pre-filtering
  - Evaluation
- Image and video quality
  - Definition of the problem
  - Applications
  - Metrics

Queen Mary University of London

Eye movements



Yarbus (1967)
- saccadic pattern

Bajcsy (1988)
- we do not see, we look

Queen Mary University of London

Video adaptation strategies

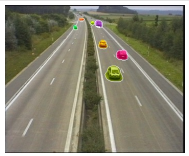
Content-blind

Coded original sequence

Spatial resolution reduction

Content-aware

Background simplification



Video enhancement

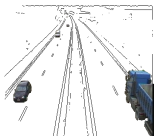
Queen Mary University of London

Background simplification

- Simplification of image areas that do not attract observer's attention (background)
  - To improve compression ratio
  - To enhance objects

Lowpass filtering

Macroblock DC value



Edge image

Uniform background

Queen Mary University of London

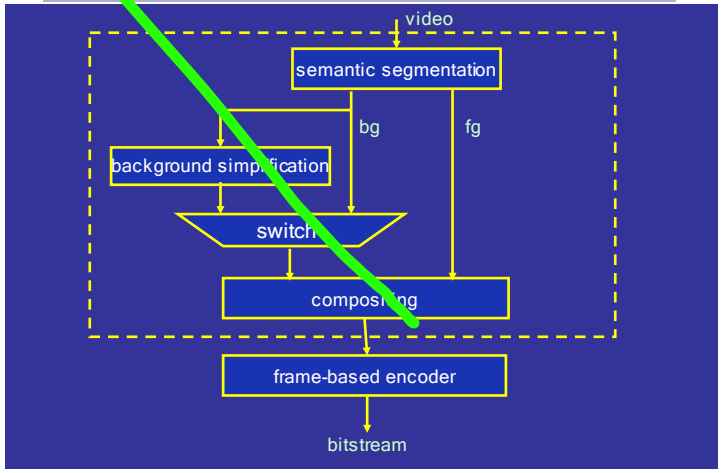
Perceptual pre-filtering

- Perceptual pre-filtering
  - decomposition of the scene into meaningful objects prior to encoding
    - foreground → region of interest
    - background → eliminated (i.e., set to a constant value) or simplified
  - takes advantage of task-oriented behavior of Human Visual System
- Objective
  - to help support low-bandwidth transmission
  - to improve perceived quality or compression ratio by
    - enhancing relevant portions of a video
    - simplifying contextual information

Queen Mary University of London

not important... （应该）不会考画这个

Perceptual pre-filtering



Subjective evaluation - details

original | Lower resolution | Fixed background | Low-pass filtered background

Queen Mary University of London

Comparison

original (CIF, 25 Hz) | 12.5 HZ | Up-sampled

simplified bg | fixed bg | low-pass filtered bg

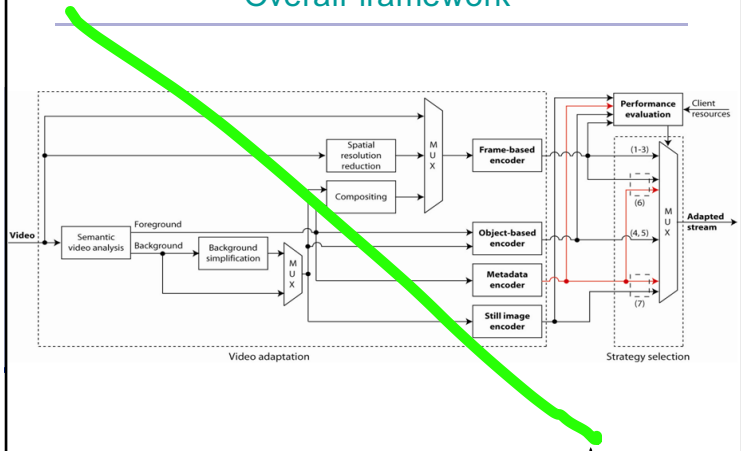
Comparison

Queen Mary University of London

Comparison

Queen Mary University of London

Overall framework

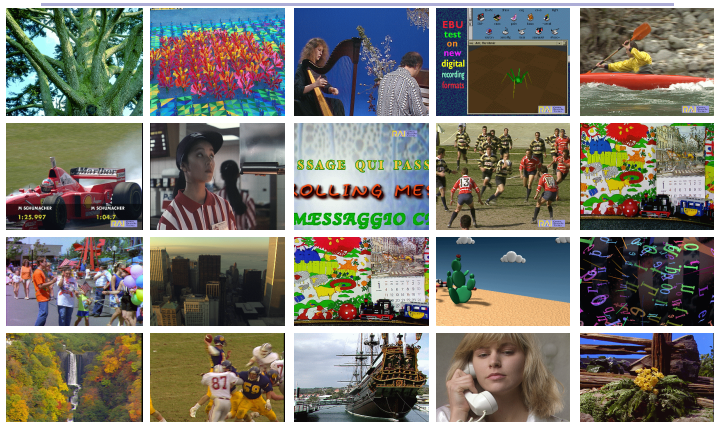


Queen Mary University of London

Today's agenda

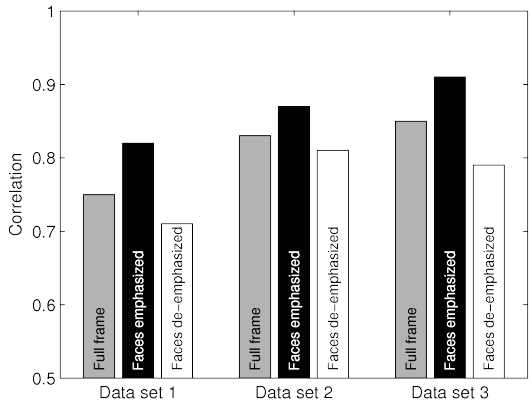
- Motivation
- Adaptive content delivery: state of the art
  - Info pyramid
  - Scalable coding
  - Transcoding
- Video analysis for content adaptation
  - Semantic pre-filtering
  - Evaluation
- Image and video quality
  - Definition of the problem
  - Applications
  - Metrics

Queen Mary University of London



Queen Mary University of London

Special case: faces



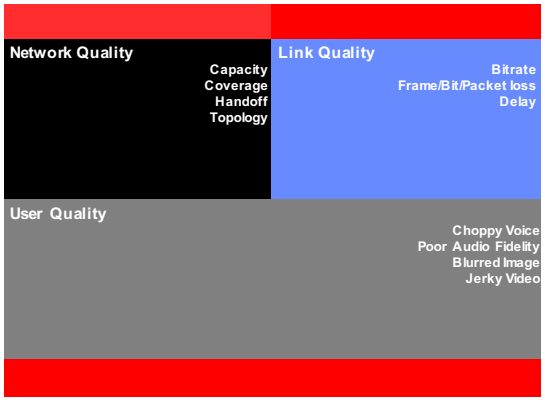
Queen Mary University of London

What is quality?

- Quality
  - One of the fundamental parameters in any system design
  - Related to the notion of distortion as a result of various signal processing
  - Key point in most optimization problems
    - If you cannot measure quality, how can you optimize it?

Queen Mary University of London

Quality framework



Queen Mary University of London

Distortions

- Pre- or post-processing
  - De-interlacing
  - Frame rate conversion
  - Quantization
  - Sampling
  - Rendering, …

Queen Mary University of London

Distortions

- Lossy compression
  - Blockiness
  - Bluriness
  - Noise
  - Jerkiness, …
- Transmission (over lossy channels) and storage
  - Transmission errors (interference, …)
  - Bit errors
  - Packet loss, …
- …

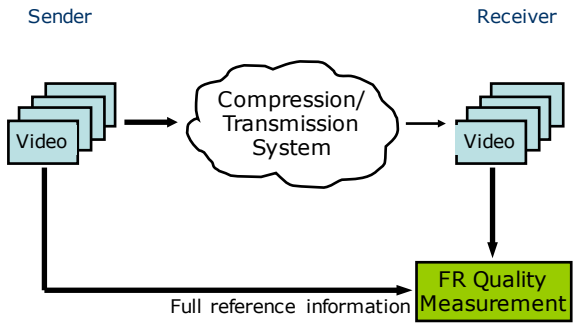
Applications

- Codec evaluation and comparison
- End-to-end testing and tuning of transmission systems
- Compression system parameter tuning
- Encoder optimization
- Restoration
- Enhancement
- Noise removal
- …

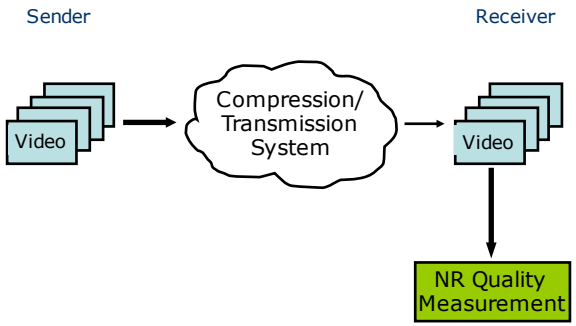
Bit-based metrics

- MSE/PSNR
  - Difference to *reference* video
  - Pixel-based
  - Mediocre quality predictors
- Network QoS
  - Bit error rate, packet loss ratio, etc.
  - Bit-/packet-based
- Limitations
  - Ignorant of content
  - Meaningless without perception

Full-reference metric



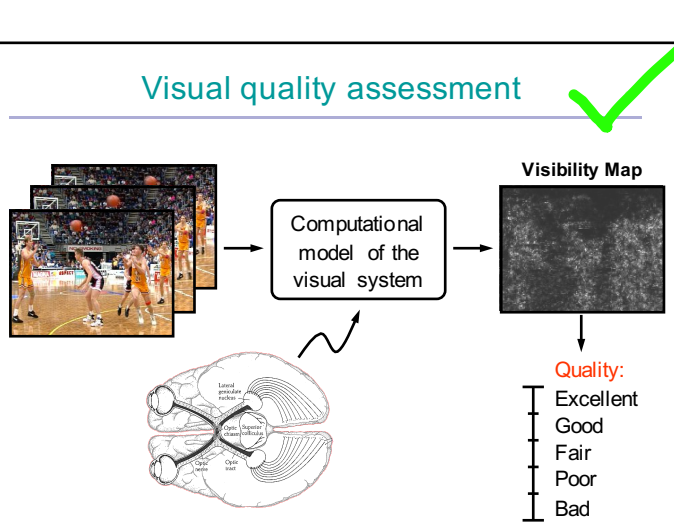
No-reference metric



Limitations …

- MSE/PSNR: believed to be a mediocre predictor
- Subjective quality assessment (Mean Opinion Score – MOS)
  - Many observers needed
  - Careful setup and control
  - Time-consuming, expensive
- Is there a way to design a more accurate objective quality metric?
  - Model of the human perception
    - Human Visual System
    - Human Auditory System
    - …

Visual quality assessment



Human visual system

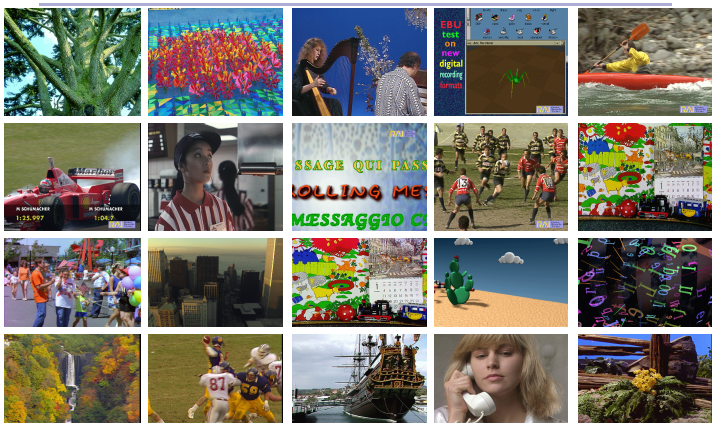
- Perceptual channels, tuned in
  - Temporal frequency
  - Spatial frequency
  - Orientation
  - Colour
- Phenomena:
  - Colour perception
  - Contrast sensitivity
  - Pattern masking
  - Integration

Quality metric evaluation

- Reference: subjective data
- Mapping of metrics' predictions to average subjective ratings
- Statistical analysis
  - Non-linear correlation
  - Rank order
  - Outliers
- VQEG (Video Quality Experts Group)
  - Evaluation of objective video quality metrics
  - Production of test sequences
  - Subjective rating experiments
  - Participants: industry, universities

Queen Mary University of London

VQEG scenes



Queen Mary University of London

VQEG Subjective tests

- Method
  - DSCQS (ITU-R Rec. 500)
  - Source, processed sequence are shown
  - Viewers rate both on a continuous scale from "bad" to "excellent" (0-100)
- Subjective testing
  - 8 labs in Europe, Japan, North America (26,000 scores recorded)
  - No universal metrics among proposals
  - YUV-PSNR performs well:
    - PSNR usually computed on uncalibrated RGB
    - Calibrated setting
    - One particular viewing setup
    - Sequence normalization
- Objective quality metrics is still an open problem

Queen Mary University of London

Quality assessment methods: summary

- Subjective quality assessment
  - Reference & benchmark
  - Standardized procedures (ITU)
  - Many observers, careful setup
  - Time-consuming, expensive
- Objective quality metrics
  - Bit-based
  - Models of the human visual system
  - Specialized artifact metrics

Queen Mary University of London

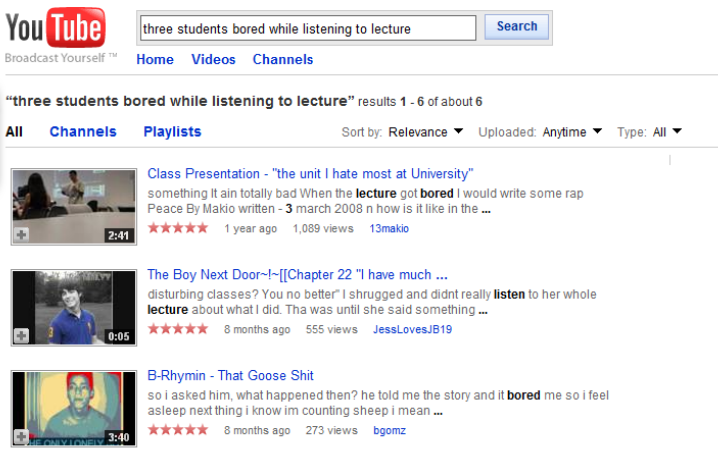
What did we learn today?

- Motivation
- Adaptive content delivery: state of the art
  - Info pyramid
  - Scalable coding
  - Transcoding
- Video analysis for content adaptation
  - Semantic pre-filtering
  - Evaluation
- Image and video quality
  - Definition of the problem
  - Applications
  - Metrics

Multimedia Systems

Multimedia content annotation

Dr. Yi-Zhe Song



EBU706U

Queen Mary University of London

Outline

- MPEG-7: multimedia content description interface
  - Metadata
  - XML description
  - Examples of applications
- Multimedia analysis for content indexing
  - Framework for adaptive delivery based on objects
  - Universal Multimedia Access
  - Content summarisation
  - Content more searchable
  - Improved visualisation
  - Event detection
  - Privacy-preserving surveillance

EBU706U

Queen Mary University of London

Metadata

semantic data

- Metadata: Data about data

raw data

- Metadata types
  - Low-level (content-based metadata)
    - low-level features extracted from the content (images, video, audio)
  - High-level (semantic metadata)
    - semantics or concepts linked to the content

EBU706U

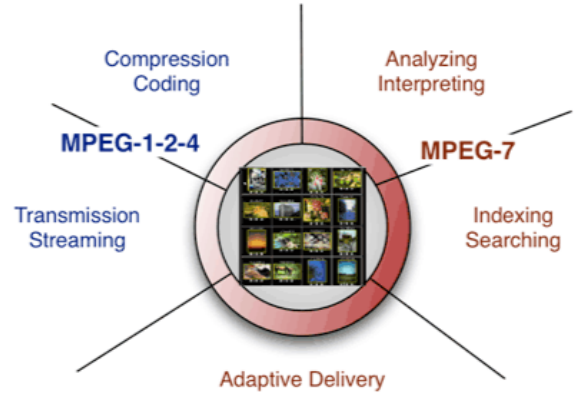
Queen Mary University of London

MPEG-7

- MPEG-7 is the international standard for "Multimedia Content Description" and it has been ratified by ISO in 2001. Unlike the more popular MPEG-2 and MPEG-4, MPEG-7 is not concerned with video compression and reproduction but with its content. MPEG-7 is not an alternative to any of the MPEG standards used for compression but complementing them and providing video with "meaning", in the form of standardized video annotations as metadata.

Queen Mary University of London

MPEG standards



Queen Mary University of London

MPEG-7: The Context

- Increasing availability of multimedia information
  - More and more people wanting to find specific multimedia information
- Difficult to get what you need:
  - searching on the web
  - selecting one of the many TV programs
  - browsing in your own home database
- Systems for searching exist on the web, but
  - mostly for textual information
  - difficult to use for 'non-engineers'
- **MPEG7:**
  - Well organised XML structure that is instantiated by the values of the features extracted from the multimedia data.
  - XML structure is easy to search, especially if it has been standardised.

Queen Mary University of London

"where MPEG7 is best used for?"

MPEG-7: applications

- Browsing and retrieval of audiovisual databases
  - image, film, radio archives
- Surveillance
  - traffic control, production chains
- Intelligent multimedia presentations
- E-commerce and Tele-shopping
  - searching for clothes / patterns
- Journalism
  - searching for events, persons
- Education, telemedicine, bio-medical applications
- Personalized television services

Queen Mary University of London

XML is the father/instantiation of HTML and MPEG7

What is XML ??

- XML stands for EXtensible Markup Language
- XML became a W3C Recommendation 10. February 1998.
- XML is a markup language much like HTML
- XML was designed to carry data, not to display data
- XML tags are not predefined. You must define your own tags
- XML is designed to be self-descriptive

- XML provides a way to structure data using a simple grammar.
- XML structures data based upon meaning, not appearance.
- XML has two components:
  - tags, used to mark the structure of the data
  - the data itself

XML & Format

- XML was created to structure, store, and transport information.
- It is just plain text. Software that can handle plain text can also handle XML.
- It is just pure information wrapped in tags. Someone must write a piece of software to send, receive or display it.
- XML-aware applications can handle the XML tags specially. The functional meaning of the tags depends on the nature of the application.

A first example (a letter):

```
<?xml version="1.0"?>

<greeting style="informal">
   <from>Chris Bates</from>
   <to>Mr. M. Mouse</to>
   <message>Hi, how are you doing?</message>
   <signature />
</greeting>
```


Queen Mary University of London

The same in HTML …

```
<html>
  <head>
      <title>The XML Sample Written in HTML</title>
  </head>
  <body>
      <h2>Chris Bates</h2>
      <h2>Mr. M. Mouse</h2>
      <p>Hi, how are you doing?</p>
   </body>
</html>
```


Queen Mary University of London

Audio Descriptor-Example

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!-- TU Berlin Audio Analyzer v1.0 http//www.nue.tu-berlin.de/forschung/projekte/mpeg7/ -->
<Mpeg7 xmlns="urnmpegmpeg7schema2001"
    xmlnsxsi="http//www.w3.org/2001/XMLSchema-instance"
    xmlnsmpeg7="urnmpegmpeg7schema2001"
    xsischemaLocation="urnmpegmpeg7schema2001 Mpeg7-2001.xsd">
<Description xsitype="ContentEntityType">
<MultimediaContent xsitype="AudioType">
<Audio xsitype="AudioSegmentType">
<AudioDescriptor xsitype="AudioSpectrumFlatnessType"  loEdge="250.0" hiEdge="600.0">
    <SeriesOfVector hopSize="PT30N1000F" vectorSize="5" totalNumOfSamples="1" >
            <Raw mpeg7dim="1 5"> 0.964695 0.967765 0.730609 0.288267 0.132297
            </Raw>
    </SeriesOfVector>
</AudioDescriptor>
</Audio>
</MultimediaContent>
</Description>
</Mpeg7>
```

Own Tags

- The tags in the example above are not defined in any XML standard. These tags are "invented" by the author of the MPEG-7 Audio Descrption DDL.
- XML language has no predefined tags.
- XML allows the author to define his own tags and his own document structure.
- The tags used in HTML (and the structure of HTML) are predefined. HTML documents can only use tags defined in the HTML standard (like <p>, <h1>, etc.).

XML image description

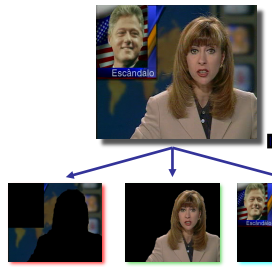
Unstructured news image

XML image description

Title → **<StillRegion id = “news”>**

</StillRegion>

XML image description



Spatial decomposition

```
<StillRegion id = “news”>
  <SegmentDecomposition
    decompositionType = “spatial”>
    <StillRegion id = “background”>
    <StillRegion id = “speaker”>
    <StillRegion id = “topic”>
  </SegmentDecomposition>
</StillRegion>
```


Queen Mary University of London

XML image description

Background feature

```
<StillRegion id = “news”>
  <SegmentDecomposition
    decompositionType = “spatial”>
    <StillRegion id = “background”>
      <DominantColor> 10 10 250
    </DominantColor>
    <StillRegion id = “speaker”>
    <StillRegion id = “topic”>
  </SegmentDecomposition>
</StillRegion>
```


Queen Mary University of London

XML image description

More features

```
<StillRegion id = “news”>
  <SegmentDecomposition
    decompositionType = “spatial”>
    <StillRegion id = “background”>
    <StillRegion id = “speaker”>
      <TextAnnotation>                    high level
        <FreeTextAnnotation> Journalist Judite Sousa
        </FreeTextAnnotation>
      </TextAnnotation>
      <SpatialMask>  the red silhouette
        <Poly>
        <CoordsI> 80 288 100 200 … 352 288 </CoordsI>
        </Poly>
      </SpatialMask>
    <StillRegion id = “topic”>
  </SegmentDecomposition>
</StillRegion>
```


Queen Mary University of London

考试题型

XML image description


```
<StillRegion id = “news”>
  <SegmentDecomposition decompositionType = “spatial”>
    <StillRegion id = “background”>
      <DominantColor> 10 10 250 </DominantColor>
    </StillRegion>
    <StillRegion id = “speaker”>
      <TextAnnotation>
        <FreeTextAnnotation> Journalist Judite Sousa
      </FreeTextAnnotation>
      </TextAnnotation>
      <SpatialMask>
        <Poly>
        <CoordsI> 5 25 10 20 15 15 10 10 5 15 </CoordsI>
        </Poly>
      </SpatialMask>
    </StillRegion>
    <StillRegion id = “topic”>
      <TextAnnotation>
        <FreeTextAnnotation> Clinton’s affair</FreeTextAnnotation>
      </TextAnnotation>
    </StillRegion>
  </SegmentDecompositio n>
</StillRegion>
```


Queen Mary University of London

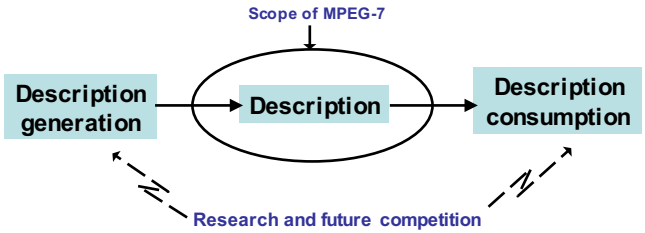
12'

MPEG-7: the objective

- Standardize a content-based description of various types of multimedia information
  - allowing quick and efficient search
  - addressing a large range of multimedia applications
  - **MPEG-1, -2, -4**: represent the content itself → 'the bits'
  - **MPEG-7**: represent information about the content → 'the bits about the bits'

Queen Mary University of London

MPEG-7: the scope



- *The description generation (feature extraction, indexing process, annotation & authoring tools,...) and consumption (search engine, filtering tool, retrieval process, browsing device, ...) are non normative parts of MPEG-7.*
- *The goal is to define the minimum that enables interoperability*

don't care if the content is correct or not

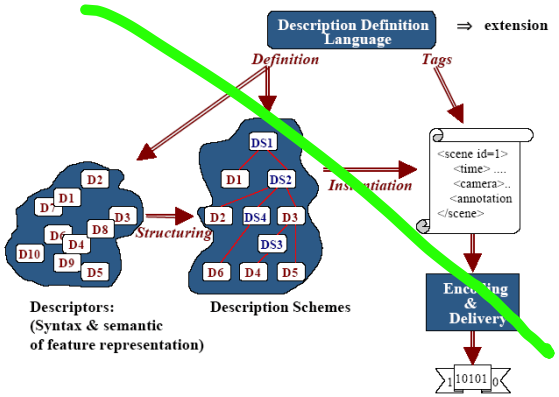
Queen Mary University of London

MPEG-7: elements

- Descriptors (**D**) to represent Features. Descriptors define the syntax and the semantics of each feature representation
- Description Schemes (**DS**): to specify the structure and semantics of the relationships between their components, which may be both Ds and DSs.
- Description Definition Language (**DDL**) to allow the creation of new DSs and, possibly, Ds and to allows the extension and modification of existing DSs.
- System tools: to support multiplexing of descriptions, synchronization of descriptions with content, transmission mechanisms, file format, etc.

Queen Mary University of London

MPEG-7: elements



Queen Mary University of London

The metadata standard: MPEG-7

- Multimedia Content Description Interface
- Part 1: Binary Format for MPEG-7 (BiM) & Terminal Architecture
- Part 2: Description Definition Language (DDL)
- Part 3: Visual Descriptors
- Part 4: Audio Descriptors
- Part 5: Multimedia Descriptors
- Part 6: Reference Software
- Part 7: Conformance
- Part 8: Extraction and use of descriptions (informative)

What is normative in MPEG-7

- Descriptors (Ds) & Description Schemes (DSs)
  - data structures
  - the feature which is represented by the data structure
- Description Definition Language (DDL)
  - XML Schema: syntax to be used for describing the data structures
  - XML: syntax to be used for the descriptions (instantiation of Ds & DSs)
- Binary Coding formats
- Profiles

- Non-normative:
  - extraction of descriptions
  - usage of descriptions

Information levels

- Low-level information
  - that is applicable to various media formats
  - for browsing of content

- Low-level descriptors for visual features
  - automatic extraction of descriptions
  - quality of descriptor (how good does the descriptor represent the feature?) can be measured by the retrieval rate
  - application independent descriptor database but
    - no idea which Ds to use to get specific semantic information
    - no idea on how to combine distance functions of descriptors (value range or function characteristic)

The Semantic Gap

- The **semantic gap** characterises the difference between two descriptions of an object by different linguistic representations, for instance languages or symbols.
- In computer science, the concept is relevant whenever ordinary human activities, observations, and tasks are transferred into a computational representation.
- More precisely the gap means the difference between ambiguous formulation of contextual knowledge in a powerful language (e.g. natural language) and its sound, reproducible and computational representation in a formal language (e.g. programming language).

The Semantic Gap

Example queries:

1) Locate any file in the known directory "/usr/local/funny".
2) Locate any file where the word "funny" appears in the filename.
3) Locate any text file where the word "funny" or the substring "humour" appears in the text.
4) Locate any mp3 file where either "funny", "comic" or "humour" appears in the metadata.
5) Locate any file of any type related to humour.
6) Locate any image that is likely to make my grandmother laugh.

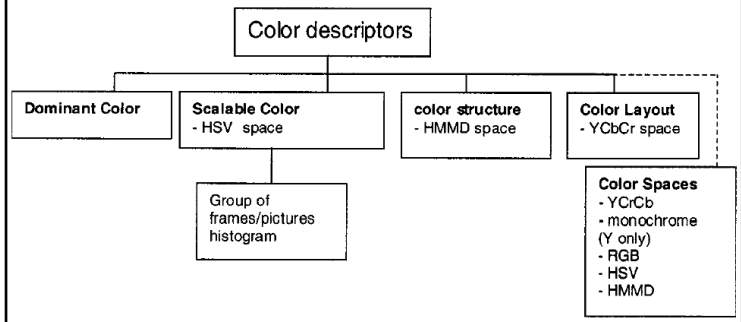
Information levels

- High-level information (semantics)
  - manual annotation of descriptions
  - very powerful description for search & retrieval (if annotation fits to application)
  - fails when annotation does not match application (does an image show a "city" ("town") or a "traffic jam"?)
  - application and semantics are defined at annotation time
  - combination of visual Ds gives semantics at search time
    - e.g. a tree has green & brown color and no edge direction at the leaves

Visual descriptors

- Basic
  - colour space
  - colour quantization
  - grid layout
- Colour
  - dominant colour
  - colour structure
  - scalable colour
  - group of frames histogram
- Texture
  - texture browsing
  - edge histogram
  - homogeneous texture
- Motion
  - motion activity
  - camera motion
  - motion trajectory
  - parametric motion
- Shape
  - contour shape
  - region shape
  - shape spectrum
  - 3D shape
  - region locator
- Faces

Colour descriptors



Scalable Color Descriptor

- A color histogram in HSV color space
- Encoded by Haar Transform

Queen Mary University of London

Dominant Color Descriptor

- Clustering colors into a small number of representative colors
- It can be defined for each object, regions, or the whole image
- $F = \{ \{c_i, p_i, v_i\}, s\}$
  - c_i : Representative colors
  - p_i : Their percentages in the region
  - v_i : Color variances
  - s : Spatial coherence

Queen Mary University of London

Color Layout Descriptor

- Clustering the image into (8x8) blocks
- Deriving the average color of each block
- Applying DCT and encoding
- Efficient for
  - Sketch-based image retrieval
  - Content Filtering using image indexing

Queen Mary University of London

Color Structure Descriptor

- Scanning the image by an 8x8 pixel block
- Counting the number of blocks containing each color
- Generating a color histogram (HMMD)
- Main usages:
  - Still image retrieval
  - Natural images retrieval

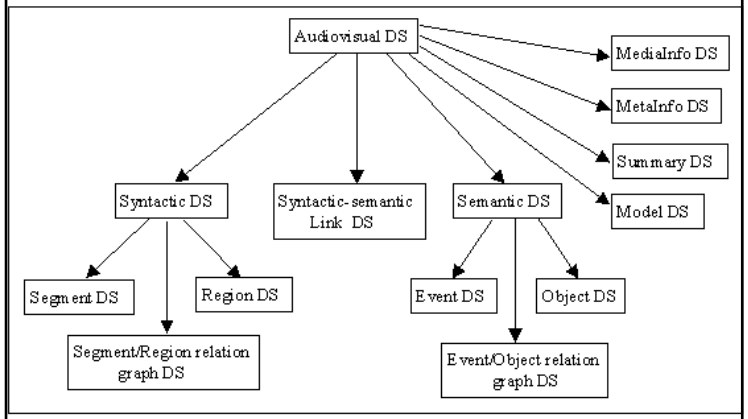
Queen Mary University of London

GoF/GoP Color Descriptor

- Extends Scalable Color Descriptor
- Generates the color histogram for a video segment or a group of pictures
- Calculation methods:
  - Average
  - Median
  - Intersection

Queen Mary University of London

Generic AudioVisual DS



Queen Mary University of London

Outline

- MPEG-7: multimedia content description interface
  - Metadata
  - XML description
  - Examples of applications
- Multimedia analysis for content indexing
  - Framework for adaptive delivery based on objects
  - Universal Multimedia Access
  - Content summarisation
  - Content more searchable
  - Improved visualisation
  - Event detection
  - Privacy-preserving surveillance

Queen Mary University of London

Automatic content annotation: why?

- Delivery to mobile devices
  - ✓ Monitoring of critical situation while shifting to intervention place
  - ✓ Remote surveillance of private grounds / vacant homes
  - ✓ Small display size
  - ✓ Restricted processing capabilities
  - ✓ Limited bandwidth
  - ✓ Time-varying conditions
- Automatic surveillance video indexing
  - Traditional video surveillance
    - Event monitoring by human observers
    - Operators located in fixed room
      - error-prone
      - tedious

Queen Mary University of London

Universal Multimedia Access

- Universal Multimedia Access
  - creation of different presentations of the same information, from a single content-base
  - adaptation of the audio-visual data to:
    - the client devices capabilities in terms of
      - computational power
      - bandwidth
    - user preferences
  - Example: **UserPreferenceDS**:
    - the user is expressing preference for a summary which contains N key frames.

Queen Mary University of London

Example: Universal Multimedia Access

What is advanced visual surveillance?

- Advanced video surveillance
  - use of automatic image and audio analysis technologies
    - motion detection
    - event detection
    - target tracking
    - face detection
    - speaker identification

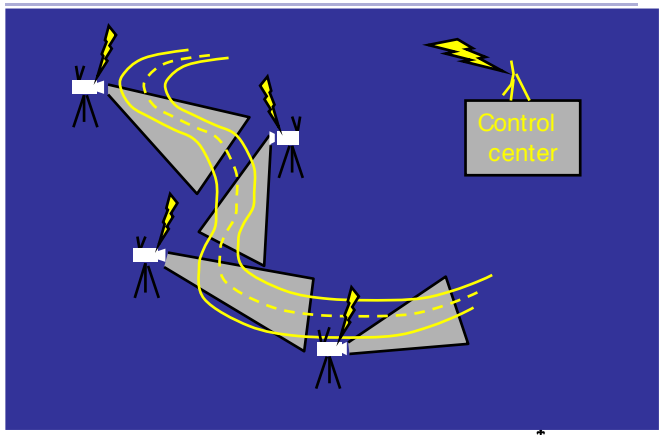
differentiate the frames
motion vector
problem: face is blurred – low resolution

  - automation of tasks that otherwise would require a human operator
- Technological factors
  - increasing processing power of low-cost microprocessors
  - decreasing cost of storage devices
  - shift from analogue to digital video cameras

Queen Mary University of London

open ended 案例分析题

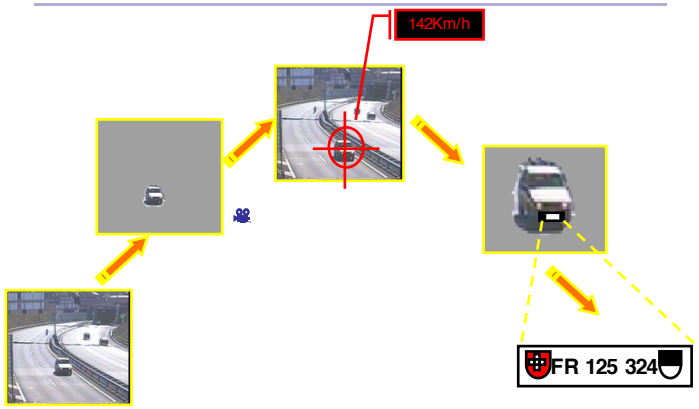
Road monitoring framework



Queen Mary University of London

1. each camera covers only a part of view
2. 'hand–off'

Traffic monitoring



Queen Mary University of London

Optical Character Recognition (OCR)

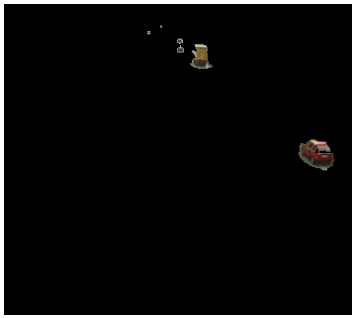
e.g. text extraction in scanned pdf

G130 WMM R846 VAR

Queen Mary University of London

Advanced surveillance

- Surveillance camera
- Object segmentation
- Description extraction
- Coding
  - Object-based (MPEG-4)
  - Text-based (MPEG-7)



Queen Mary University of London

MPEG-7 camera

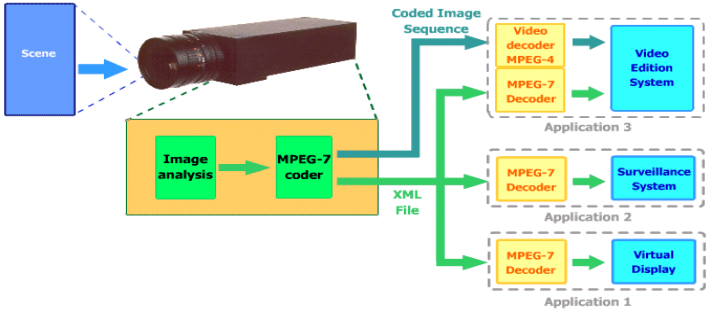
- MPEG-7 camera
  - describes a scene in terms of semantic objects and of their properties
  - use semantic video analysis (including tracking)
  - extract and transmit only the relevant information (object motion)

XML scene description

Queen Mary University of London

Intelligent camera

- MPEG-7 Camera
  - Image analysis block: includes image analysis algorithms (implemented on the camera DSP) such as segmentation, change detection, and tracking.
  - MPEG-7 coder: scene description represented using MPEG-7 (XML).
  - MPEG-7 decoder: MPEG-7 description is parsed. Extraction of the information related to the specific applications.

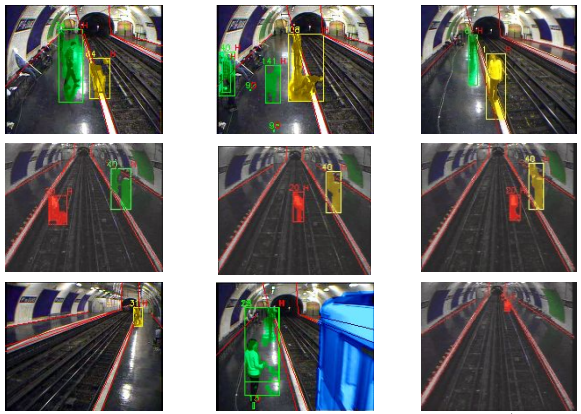


XML scene description

```
<Object id="4">
  <RegionLocator>
    <BoxPoly> Poly </BoxPoly>
    <Coords1> 237 222 </Coords1>
    <Coords2> 230 252 </Coords2>
    <Coords3> 240 286 </Coords3>
    <Coords4> 308 287 </Coords4>
    <Coords5> 312 284 </Coords5>
  </RegionLocator>

  <DominantColor>
    <ColorSpace>  YUV  </ColorSpace>
    <ColorValue1> 143.4 </ColorValue1>
    <ColorValue2> 123.3 </ColorValue2>
    <ColorValue3> 128.2 </ColorValue3>
  </DominantColor>

  <HomogeneousTexture>
    <TextureValue> 0.02 </TextureValue>
  </HomogeneousTexture>

  <MotionTrajectory>
    <TemporalInterpolation>
      <KeyFrame>  100  </KeyFrame>
      <KeyPos> 268.6 251.7 </KeyPos>
      <KeyFrame>  101  </KeyFrame>
      <KeyPos> 262.8 241.0 </KeyPos>
                ...
      <KeyFrame>  138  </KeyFrame>
      <KeyPos> 192.9 79.0 </KeyPos>
    </TemporalInterpolation>
  </MotionTrajectory>

</Object>
```

badly–lit

Event detection in multi-camera networks



Queen Mary University of London

CCTV

- How many times a day are you caught on CCTV cameras?

- The average citizen is caught on CCTV cameras 300 times a day
- There are more than 30 million CCTV cameras in operation worldwide
- Video surveillance market is large and growing!

Queen Mary University of London

Privacy: current status

- 'Old' CCTV: analogue video
  - videotape systems
  - forensic: slow search
- Current and future CCTV: digital video
  - images stored on DVDs or CDs
  - can be indexed and searched easily
  - can be quickly transmitted over networks
- Issue
  - with the development of automation and technology, advanced video surveillance may generate an undesired effect: people might not perceive it as a security tool but as a threat to their private sphere

Queen Mary University of London

MPEG-7 camera applications

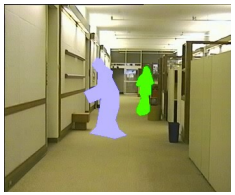
- Virtual display
  - The virtual objects (e.g., blobs) follow the movement of the persons
  - Applications
    - Privacy: only the behavior of the persons are transmitted
    - Checking intentions in surveillance
    - Extract various statistics without revealing identity of people

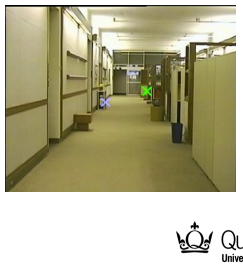
real scene moving objects virtual objects

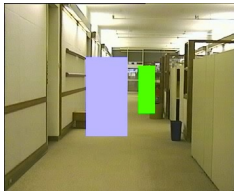
Queen Mary University of London

Privacy-preserving surveillance

rendering covering identities



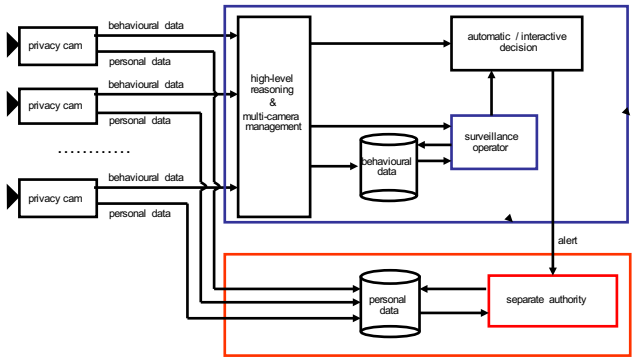




Queen Mary University of London

跳了跳了

Privacy-preserving surveillance system



Queen Mary University of London

What did we learn today?

- MPEG-7: multimedia content description interface
  - Metadata
  - XML description
  - Examples of applications
- Multimedia analysis for content indexing
  - Framework for adaptive delivery based on objects
  - Universal Multimedia Access
  - Content summarisation
  - Content more searchable
  - Improved visualisation
  - Event detection
  - Privacy-preserving surveillance

Past exam questions

1. What is MPEG7? Explain the difference between MPEG4 and MPEG7 and give five relevant applications of MPEG7
2. MPEG7 standardizes several low-level descriptors for image and video. Give at least three different MPEG7 visual descriptors related to each one of the following primitives: colour, texture, shape and motion.
3. The semantic gap is a current limitation for content-based multimedia retrieval applications. Explain the difference between low-level multimedia metadata and semantic metadata.
4. The MPEG-7 Framework (standard) consists of normative and non-normative parts. Outline the four normative and the two non-normative structures in this framework.
5. Describe the privacy issue in audio-visual surveillance and explain how technology can be used to mitigate this problem.