A Framework for Online Semantic Adaptation of Scalable Video

Luis Herranz
Grupo de Tratamiento de Imágenes
Universidad Autónoma de Madrid
December 2006

Outline

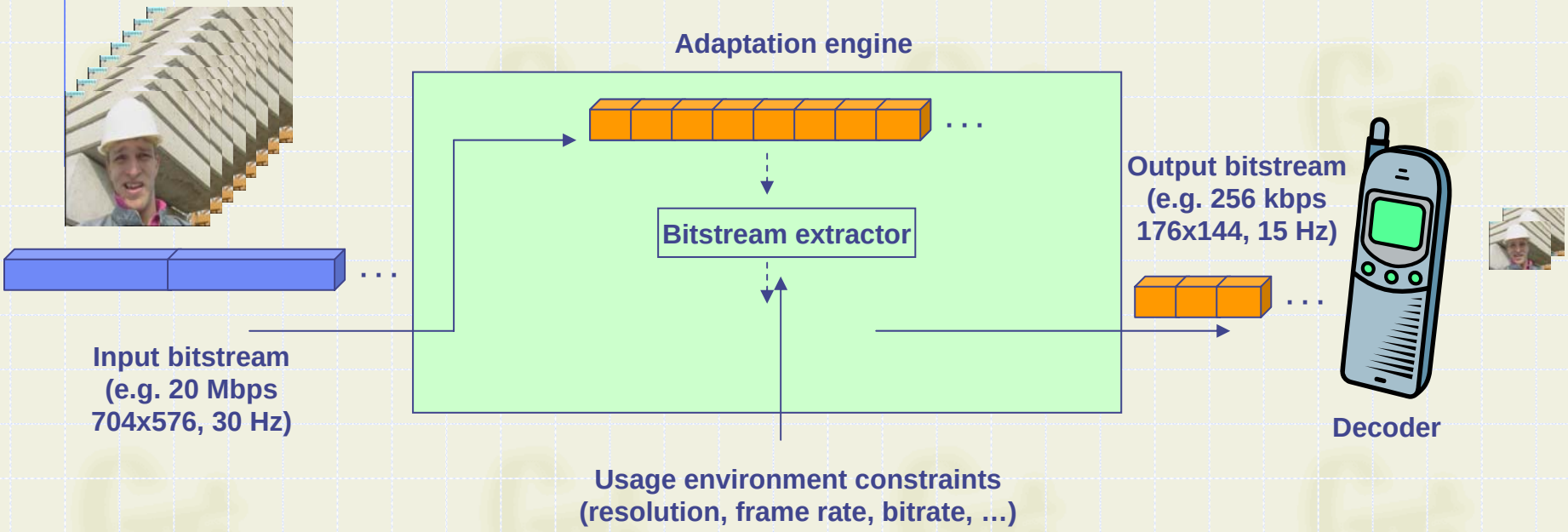
- Introduction (motivation)
  - Scalable video adaptation
  - Video adaptation using semantic information
- Proposed framework
  - Scalable video adaptation using semantic clues
- Activity based skimming
  - Activity analysis
  - Including usage environment constraints
- Conclusions

Introduction. Scalable video adaptation (I)

- Scalability
  - A *single* bitstream contains *multiple* versions of the same content
  - Allows to adapt a content without need of decoding and reencoding
- Fully scalable video codec
    - Temporal scalability: frame rate (less frames per second)
    - Spatial scalability: resolution (less pixels)
    - Quality (SNR) scalability: bitrate
  - Bitstream parts (atoms) can be arranged in a cube structure
- Wavelets, H.264 based (e.g. MPEG-4 SVC), ...
  - In this work is used a fully scalable video codec based on wavelets (developed by QMUL)

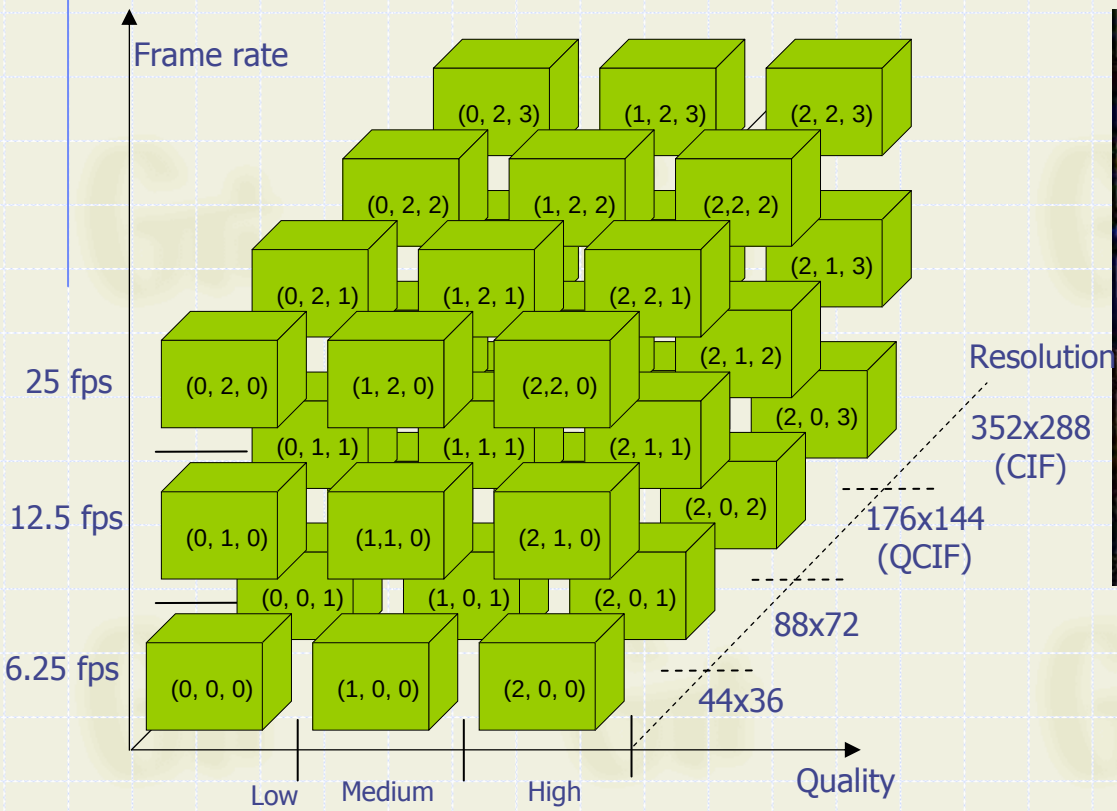
Introduction. Scalable video adaptation (II)

- Adaptation
  - Simply bitstream extraction
    - Very efficient



Introduction. Scalable video adaptation (III)

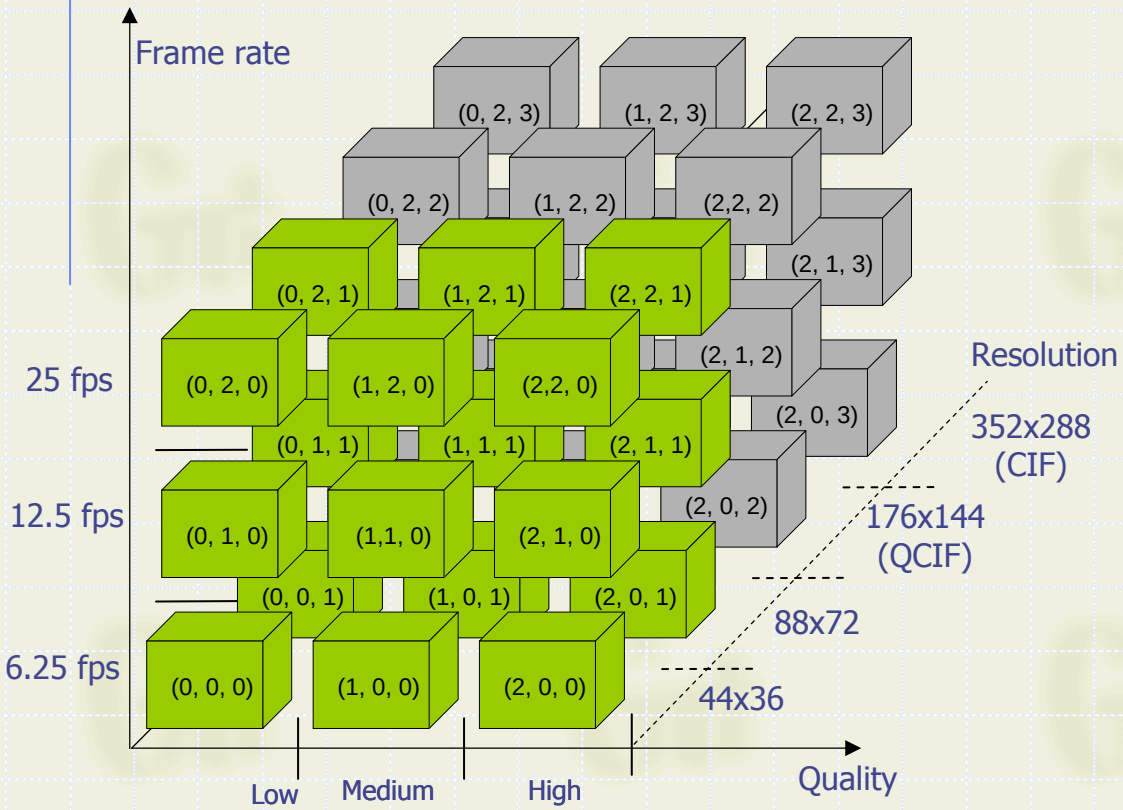
- Cube structure
  - Adaptation unit: Group of frames (GoF)



352x288, 25fps, High bitrate

Introduction. Scalable video adaptation (III)

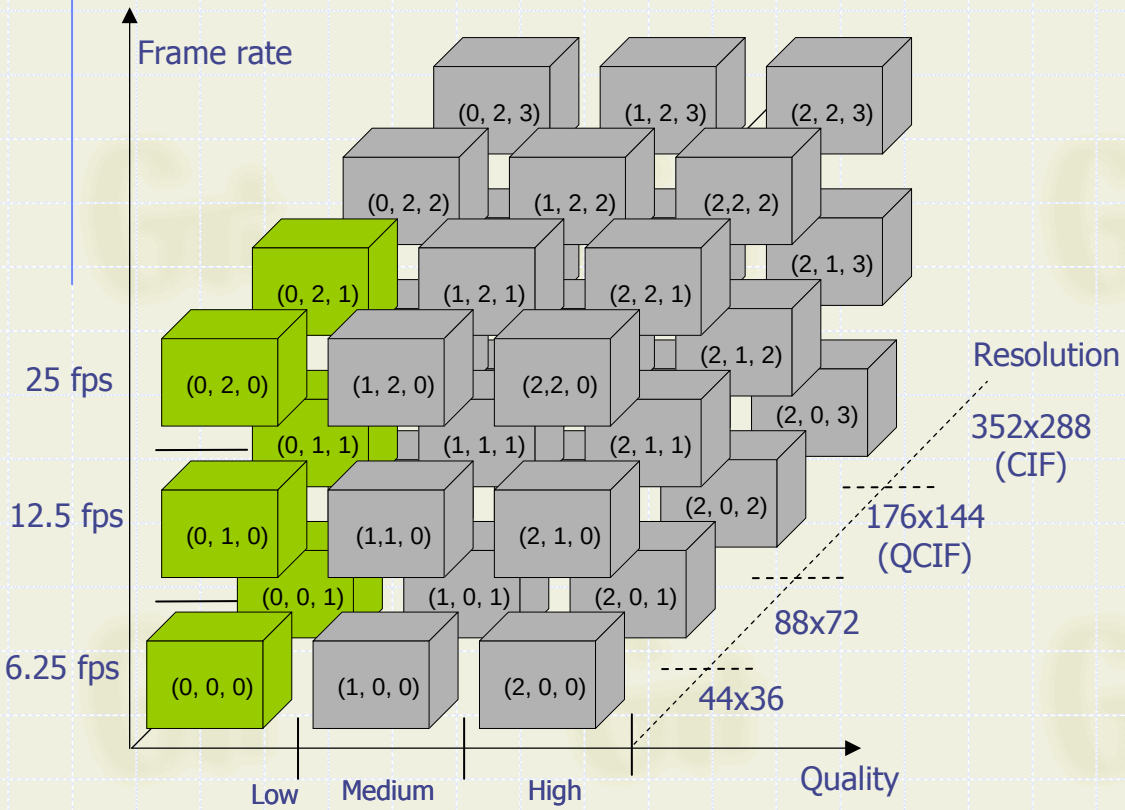
- Cube structure
  - Adaptation unit: Group of frames (GoF)



88x72, 25fps, High bitrate

Introduction. Scalable video adaptation (III)

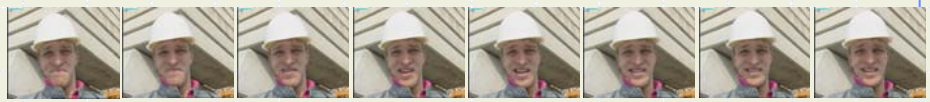
- Cube structure
  - Adaptation unit: Group of frames (GoF)

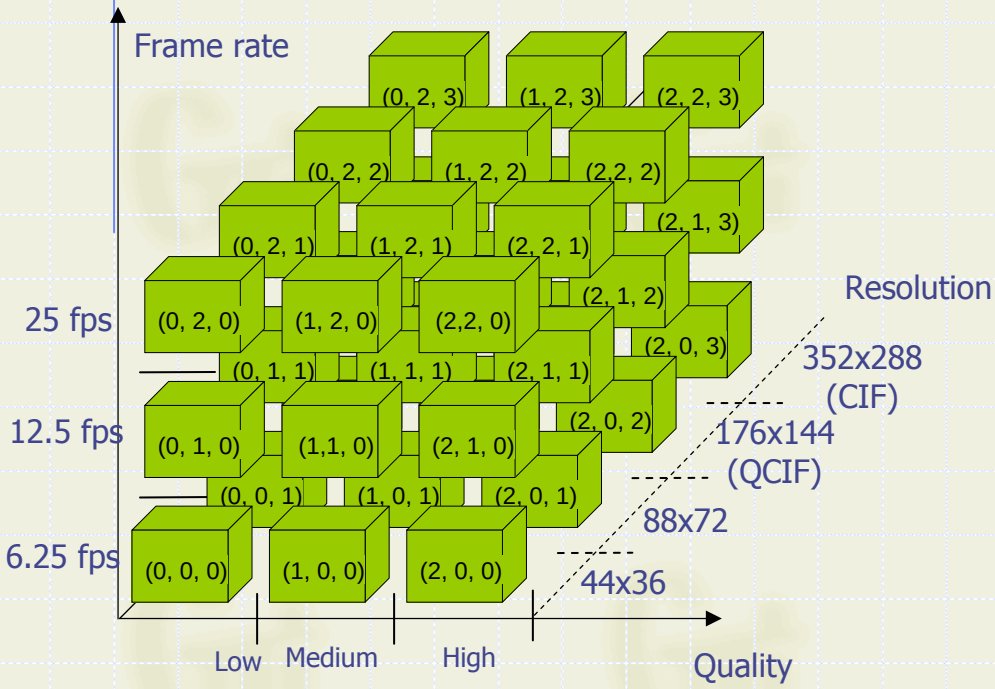


88x72, 25fps, **Low bitrate**

Introduction. Scalable video adaptation (III)

- Cube structure
  - GoF length 8 frames

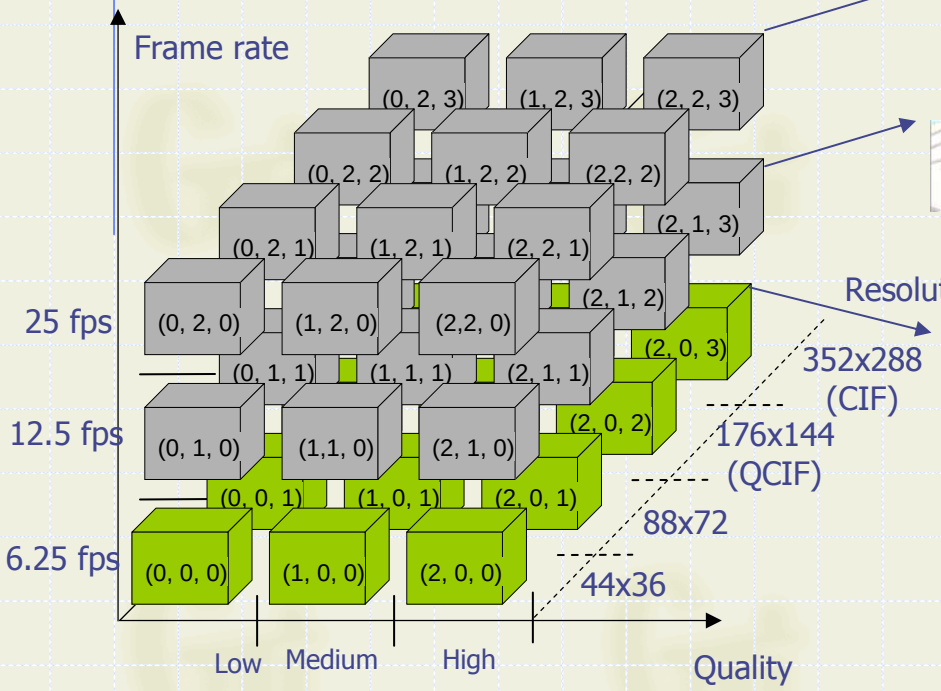




352x288, **25fps**, High bitrate

Introduction. Scalable video adaptation (III)

- Cube structure
  - GoF length 8 frames



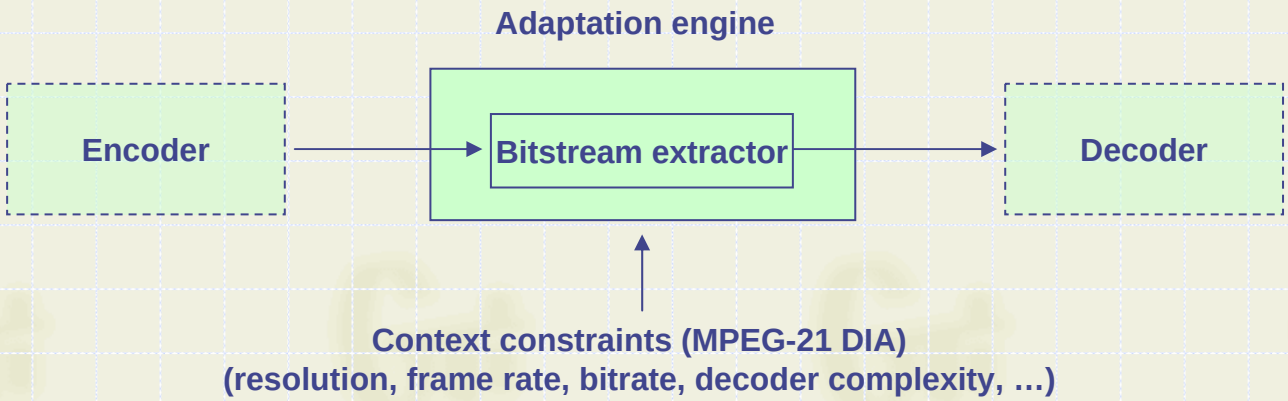
352x288, **6.25fps**, High bitrate

Introduction. Semantic video adaptation

- Video adaptation using semantic information
  - Usually in adaptation we lose some information
    - Try to preserve those parts more "semantically" relevant
  - Requires analysis of the content
    - Previously stored (metadata)
    - Online. Usually from the compressed domain
  - Examples of semantic adaptation
    - Video summarization
    - Video skimming
    - "Smart" fast browsing/playback
  - Online adaptation
    - Adaptation as the bitstream arrives

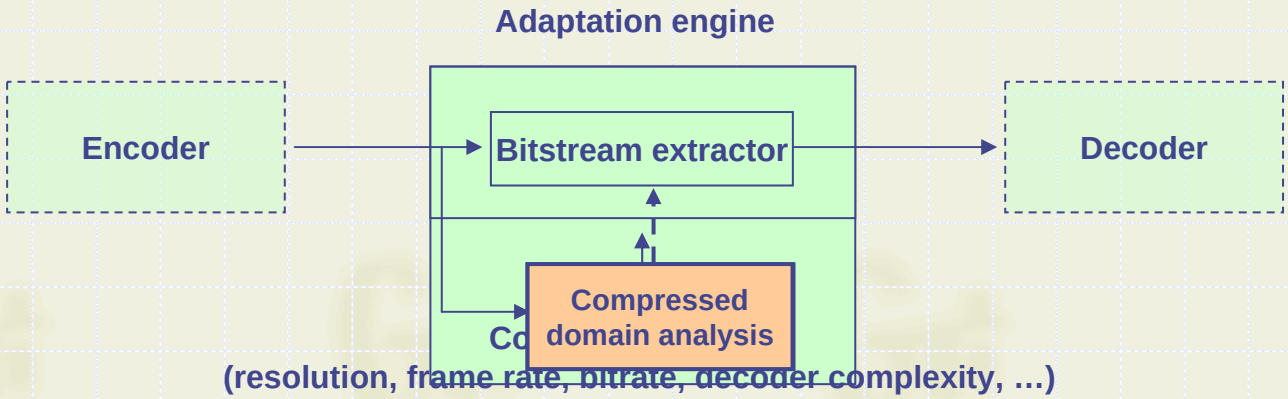
Proposed framework (I)

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



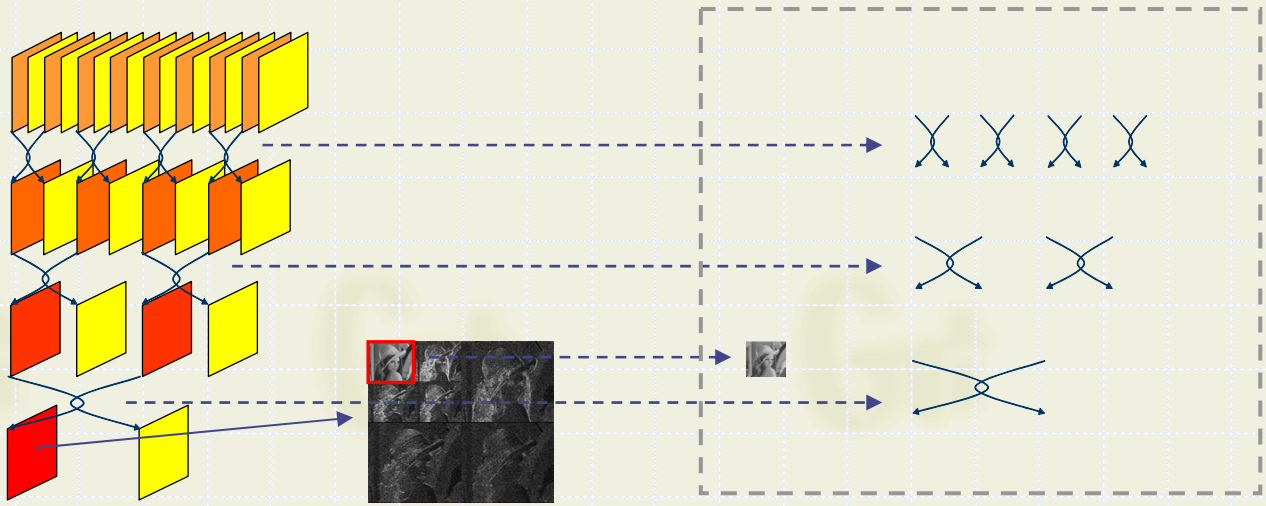
Proposed framework (II)

- Semantic adaptation engine for scalable video
  - Context aware (MPEG-21 Usage enviroment descriptions)
  - Content aware
    - Bitstream extraction guided by content (analysis)
  - Efficiency relies on efficient analysis
    - Compressed domain analysis
    - Shot change detection, activity, …



Proposed framework (III)

- 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

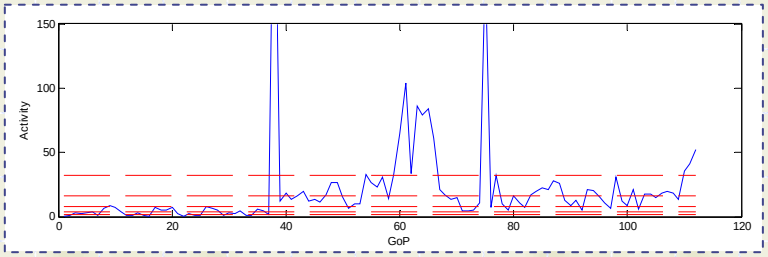


Activity based skimming (I)

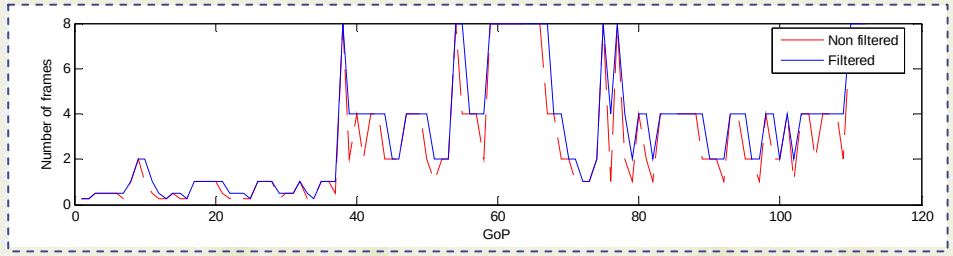
- As an application of the previous framework
- Dynamic selection of the temporal level
  - GoF by GoF (online)
  - According to GoF activity analysis
  - Adaptive frame rate

Activity based skimming (II)

- 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



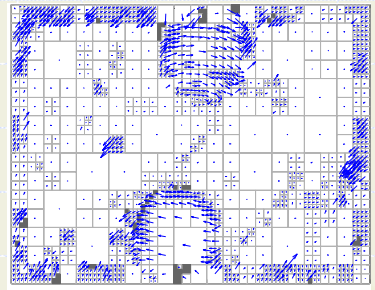
GoF activity



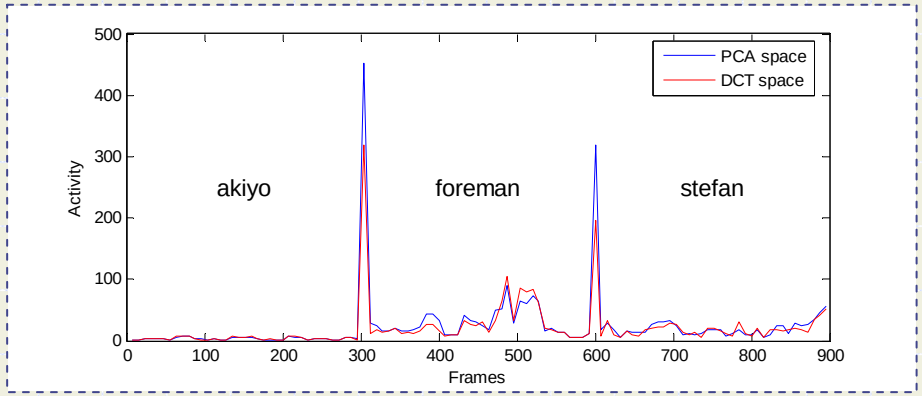
Skimming curve

Activity based skimming (III)

Experiments on GoF activity

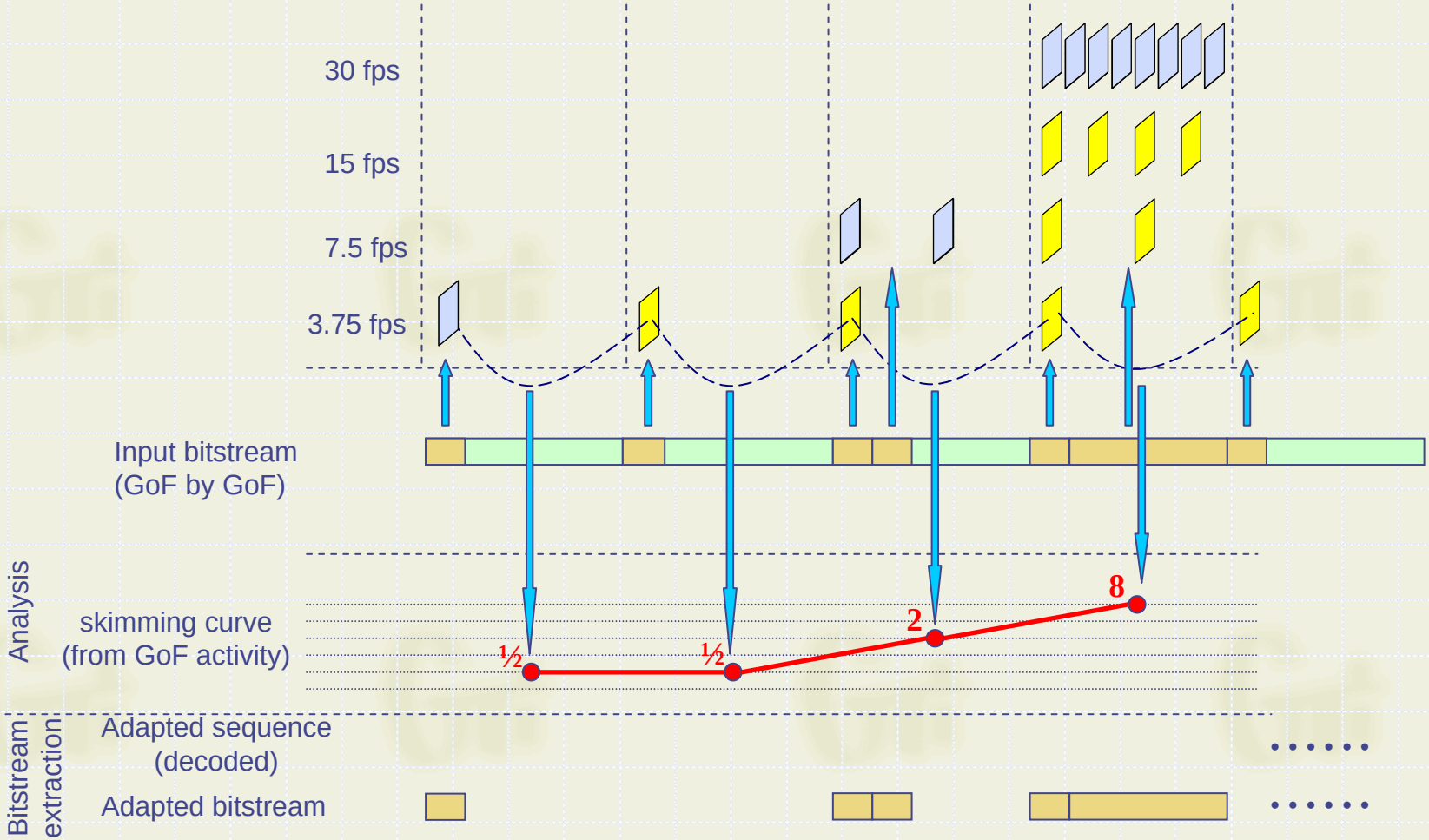


- Using motion vectors
  - MPEG-7 intensity of motion activity
    - Problems: highly dependent on the codec implementation
    - Sometimes is still computationally expensive
- Using low resolution versions
  - Lowest frame rate, lowest frame size : 1 frame per GoF
  - PCA analysis was proposed as an 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



Activity based skimming (IV)

- Semantic activity-based adaptation on scalable video



Including the usage environment (I)

- 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*

Including the usage environment (II)

- 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*)

Including the usage environment (III)

Results

Source sequence
(352x288 30 fps)

Adapted sequence
(176x144 fast browsing at 30 fps)

Adapted sequence
(176x144 dynamic frame rate decoder)

Summary

- 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 activity based video skimming
  - Experiments on activity analysis over scalable video
  - Integrated into the standard MPEG-21 DIA framework

Thank you