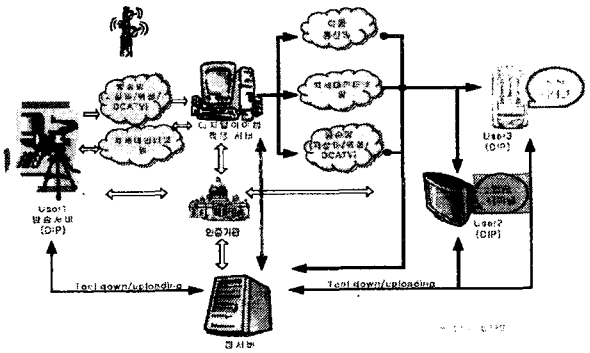
방송통신 융합 서비스를 위한 컨텐츠 적응 기술
(Contents Adaptation for the Convergence Service of Broadcasting and Telecommunication)

2004. 10. 13

Yong Man Ro
ICU
yro@icu.ac.kr

Convergence of Broadcasting and Telecommunication

- Regulatory boundaries between communications and broadcasting become blurred.
  - Media service in mobile terminal, Internet Broadcasting, VOD over ADSL, Internet over CATV, DMB, Data broadcasting
- The convergence
  - Network
  - Contents
    - User Preference, Environment(Anytime/Anywhere), Terminal adaptation,

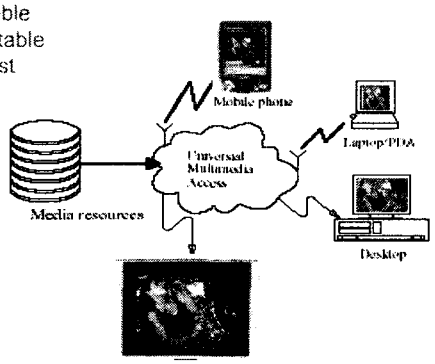


Background

- Complex ecosystem between media creator, provider, rights manager, delivery infrastructure, media consumer
  - Mostly non-transparent & proprietary information exchange
  - Heterogeneous ***Multimedia Applications/Services***
    - interactive/non-interactive, realtime/nonrealtime,low delay/high bandwidth, etc.)
  - Heterogeneous ***Devices***
    - screen sizes, CPUs, memory, power supplies, interfaces, etc.
  - Heterogeneous ***Access Networks***
    - Varying characteristics for loss, bandwidth, reliability, etc
  - Heterogeneous ***User Policies***
- *Need a **multimedia framework** to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities*

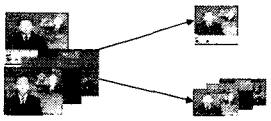
Background

- Universal Multimedia Access (UMA)
  - Universal Multimedia Access (UMA) deals with delivery of images, video, audio and multimedia content under different network conditions, user and publisher preferences, and capabilities of terminal devices.
- Media communications will be
  - Scalable
  - Adaptable
  - Robust



Content Adaptation

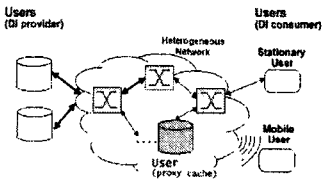
- Contents Requirement for the Convergence
  - Reuse /Copy One Publish Many
  - Meet heterogeneous environment(terminal, network)
  - Best quality of experience
- Talk in this seminar...
  1. Contents adaptation (MPEG-21 DIA)
  2. Contents adaptation and Media QoS
  3. Transcoding for Media QoS
  4. Modality Conversion
  5. Demos



1. Contents Adaptation in MPEG-21 (MPEG-21 DIA)

MPEG-21 Standardization for Multimedia Framework

- Goal
  - Define the technology needed to support *Users* to exchange, access, consume, trade, and otherwise manipulate *Digital Items* in an efficient, transparent and interoperable way
  - Define the relation between main elements of framework for multimedia transaction
- Essential concepts : *Digital Item, User*
  - *Digital Item* : the "*What*" of the multimedia Framework
  - *User* : the "*Who*" of the multimedia Framework



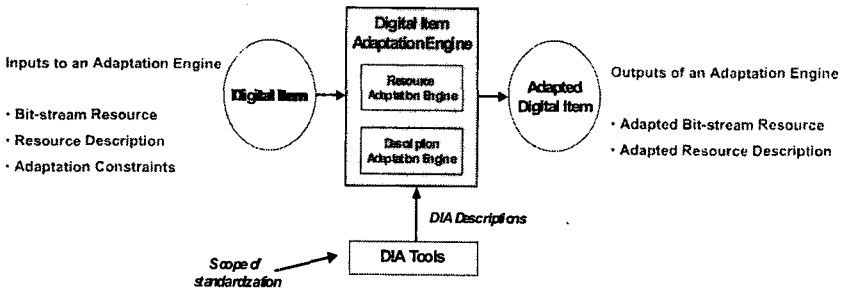
<Example Distributed Multimedia System, source: MPEG>

MPEG-21 Digital Item Adaptation(DIA)

- Background
  - Diversity of networks and terminals is a formidable obstacle to universal media access
    - bandwidth, display/audio capabilities, processing power, user preference
    - Adaptation necessary in the delivery route
    - Need standardization of metadata interfacing to adaptation engines to open up adaptation services
  - Diversity of formats
    - Currently only a few predominant formats
    - In the future there will be more
      - Standardized and Proprietary, combinations, rich media
      - Even the content type may be unknown
    - Can adaptation be made fully format agnostic?
- Goal:
  - To define description tools to enable and enhance the adaptation of Digital Item
  - To achieve transparent interoperable access to distributed multimedia content

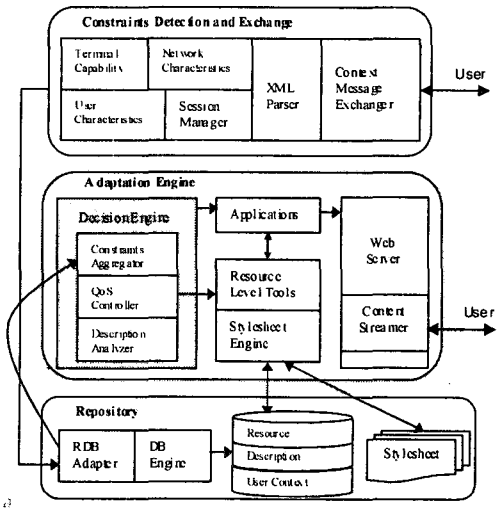
MPEG-21 Part 7: DIA

- *Descriptions and format-independent mechanisms* in terms of resource adaptation, description adaptation, and/or Quality of Service management *are within the scope of the standardization*



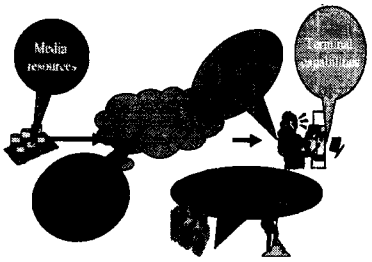
<Concept of MPEG-21 DIA>

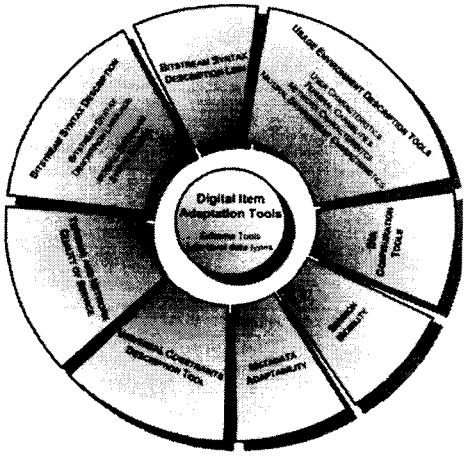
Content Adaptation System



MPEG-21 DIA Tools

Clustering according to their functionality and use for Digital Item Adaptation





Overview and Organisation of Digital Item Adaptation Tools

DIA Tools overview

Usage Environment Description Tools

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

Digital Item Resource Adaptation Tools

- Bitstream Syntax Description
- Terminal and Network QoS
- Metadata Adaptability

Digital Item Declaration Adaptation Tools

- Session Mobility
- DID Configuration Preferences
- DIA Description Messages

User Environment Description tools

User Characteristics
Specifies tools for describing various characteristics of users:
- General user information
- Usage preferences and history
- Presentation preferences
- Accessibility characteristics
- Mobility characteristics and destination

Terminal capabilities
To satisfy consumption and processing constraints of a particular terminal, consists of:
- Codec capabilities
- Device properties (power, storage and data I/O characteristics
- Input-output characteristics

Network Characteristics
Used for efficient and robust transmission and resources:
- Network capabilities and conditions (Available bandwidth, delay and error characteristics)

Natural Environment Characteristics
Specifies tools for describing natural environment characteristics including :
- Location and time of usage of a Digital Item
- Characteristics that pertain to audio-visual aspects

UED: User Characteristics

- User , UserInfo, UsagePreferences, UsageHistory , Destination
- AudioPresentationPreferences (e.g., `VolumeControl`)
- ColorPreference (e.g, `ColorTemperaturePreference`)
- StereoscopicVideoConversion, DisplayPresentationPreferences
- GraphicsPresentationPreferences (eg, preferred emphasis of geometry)
- FocusOfAttention
- PresentationPriorityPreference (e.g. `ModalityPriorities`)
- ConversionPreference
  - conversions may be carried out when a terminal or network cannot support a particular modality or format
- VisualImpairment
  - The low vision conditions described by the User's symptoms,
- AuditoryImpairment
  - The hearing threshold shift of a User is described
- ColorVisionDeficiency
  - to deliver recognizable color with color vision deficiency
- MobilityCharacteristics,
  - the movement of a User (directivity, location update intervals,erraticity

UED: Terminal Capabilities

- Specifies encoding and decoding capabilities, device properties, which include power, storage and data I/O characteristics, input-output characteristics , display and audio output capabilities
- Terminal, CodecCapability, CodecCapability, CodecParameter
  - encoding or decoding
- Displays, Display, Display capability
  - resolution, color capabilities and/or rendering format
- AudioOutput, AudioOutputCapabilities
  - This is achieved by specifying the sampling frequency and bits-per-sample, the frequency range of the output, the number of channels supported, as well as power and signal-to-noise ratio.
- UserInteractionInput ,DeviceClass
- PowerCharacteristics, Storage, StorageCharacteristics, DataIO, DataIOCharacteristics,
- Benchmark, CPUBenchmark, ThreeDBenchmark

UED:Network Characteristics

- Specifies the characteristics of multiple networks, where the characteristics of a single network are specified by the Network element in the NetworksType
- Network, NetworkCapability,
  - maximum capacity of a network and the minimum guaranteed bandwidth that a network can provide. Also specified are attributes that indicate if the network can provide in-sequence packet delivery and how the network deals with erroneous packets
- NetworkCondition
  - available bandwidth, error and delay. The error is specified in terms of packet loss rate and bit error rate. Several types of delay are considered, including one-way and two-way packet delay, as well as delay variation. Available bandwidth includes attributes that describe the minimum, maximum, and average available bandwidth of a network.

UED: Natural Environment Characteristics

- Specifies the characteristics of multiple natural environments, where the characteristics of a single natural environment are specified by the NaturalEnvironment element in the NaturalEnvironmentsType
- NaturalEnvironment
  - Specific characteristics include the location and time of usage of a Digital Item, as well as audio-visual characteristics of the natural usage environment.
- Location, Time
- AudioEnvironment
  - describes the natural audio environment of a particular User in terms of the measured noise level and noise frequency spectrum.
- IlluminationCharacteristics
  - These attributes are related to the effect of illumination on the perceived color of a displayed image and can be used to control the color so as to present the original color that is intended.

DI Resource Adaptation Tools

Terminal and Network QoS

Addresses the problem of selecting optimal parameter settings for media resource adaptation to satisfy constraints imposed by terminals and/or networks while maximizing the QoS, Specifies the relationship between:

- Constraints, feasible adaptation operations satisfying these constraints
- Possibly associated utilities and QoS management achieved by adaptation of media resources to imposed constraints

Bit-Stream Syntax Description

The resulting XML document to describe the high-level structure of a bitstream:

- Acts as additional layer, similar to metadata (address its high level structure-how it is organized in layers or packets)
- Scalable: describe BS at different syntactic layers
- Adaptable to properly reflect bitstream adaptations

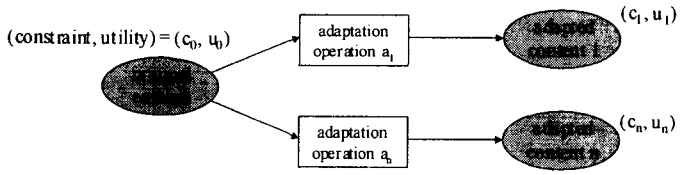
Metadata Adaptability

Describing adaptation information pertaining to metadata within a Digital Item, a set of syntactical elements with prior knowledge about the metadata that is useful for reducing the complexity of the metadata adaptation process, to enable

- Filtering and scaling of descriptions
- Integration of descriptions

Terminal and network QoS

- The `AdaptationQoS` descriptor specifies the relationship between constraints, feasible adaptation operations satisfying these constraints, and associated utilities (qualities) => allowing the selection of optimal adaptation parameters.



Example:

Given the limited constraint c* from terminal/network, which adapted content (corresponding to a specific adaptation operation) will be selected to present to user?

The solution is: select (c_i, u_i, a_i) so that $c_i < c^*$ and u_i is the highest possible.

Prof Y M. Ro in IVY-Lab

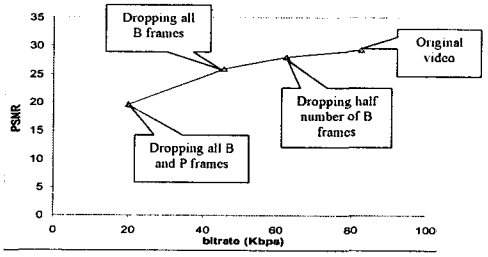
Terminal and network QoS

- AdaptationQoS has 3 separate modules:
  - UtilityFunction provides a restricted set of adaptation operation points in a list format to choose from.
  - LookUpTable is a matrix representation format, enabling selection by interpolation of an adaptation operation point and allowing extra info to be represented.
  - StackFunction is a functional representation format.

Utility function (Terminal and network QoS)

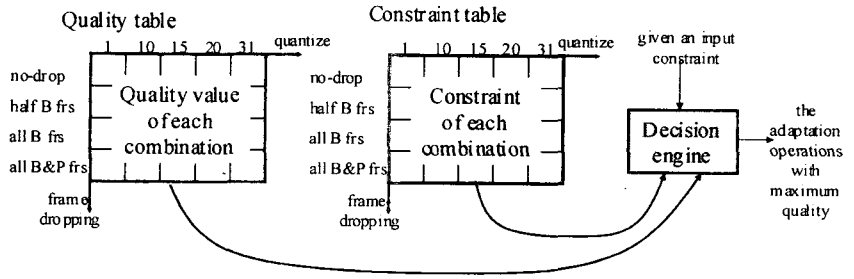
- UtilityFunction provides a restricted set of adaptation operation points in a list format. Each point has three elements:
  - Constraint (a set of sample points of interest in the constraint space)
  - AdaptationOperator (permissible adaptation operations for the constraint)
  - Utility (the value of video utility after the adaptation by AdaptationOperator.
- Example:

list of 4 points
- Constraint (KBps): 19, 45, 64, 85
- Adaptation operators: no-drop, drop half B frs., drop all B frs., drop all B & P frs.
- Utility (PSNR): 29.1, 28.2, 25.9, 19.8



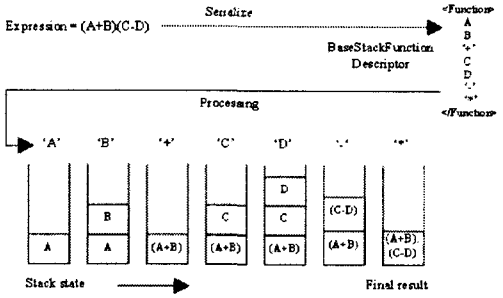
LookUpTable (Terminal and network QoS)

- In some cases, the relationship among Constraint, AdaptationOperator, Utility is too complex to put in a list format.
- This relationship can be put into matrix format (table)
- ***Example:***



StackFunction (Terminal and network QoS)

- The UtilityFunction and LookUpTable show the relationship of different data in a enumerative way (listing all possible cases)
- Sometimes, the provider wants to represent that relationship by analytical functions => using the StackFunction
- Note: the analytical functions can be obtained by fitting the function to the enumerative data of UtilityFunction or LookUpTable
- BaseStackFunction mechanism:



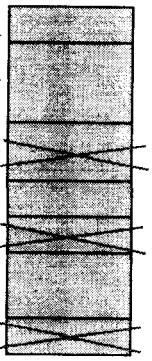
Dynamic Content Adaptation

- How to fully exploit scalability for content adaptation?
- In a client-server application:
  - The client requests a scalable content available on a server
  - The server establishes a content negotiation stage
  - The server edits the bitstream, and sends the bitstream
- Need of a software to parse and edit the bitstream
  - each coding format has its own data structure

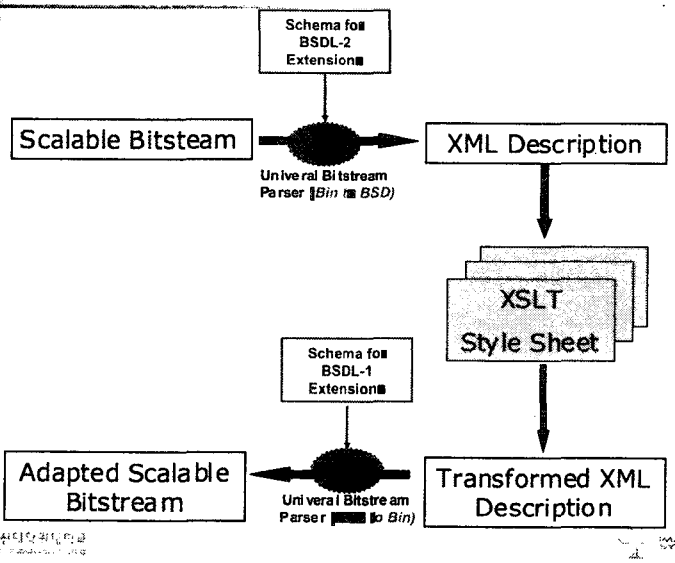
Generic approach: use a common structuring language (XML) to describe the bitstream syntax (Bitstream Syntax Description (BSD))

BSD: XML description

<Bitstream
xml:base="http://www.example.com/akiyo.mpg4>
<Header> 0-17 </Header>
<frame> 18-4658 </frame>
~~<frame> 4659-4756 </frame>~~
<frame> 4757-44772 </frame>
~~<frame> 4773-4758 </frame>~~
<frame> 4760-4973 </frame>
~~<frame> 4974-5026 </frame>~~
</Bitsteam>



BSD: Edition of a multimedia bitstream with XSLT

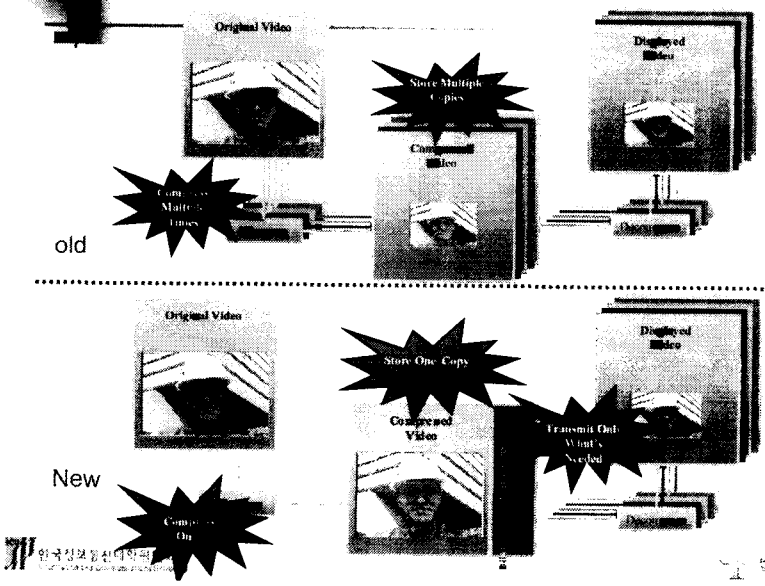


2. Contents Adaptation and Media QoS

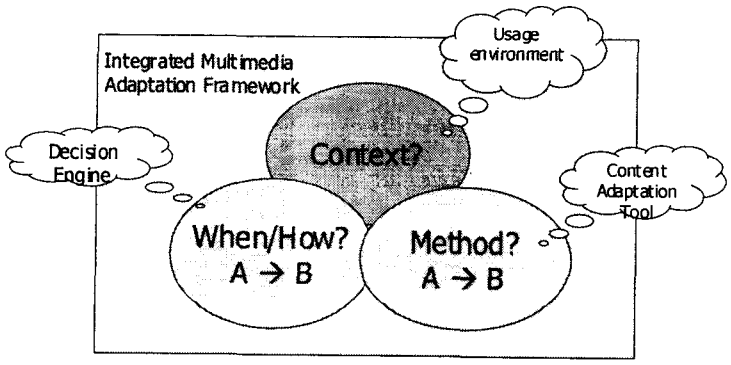
Contents Adaptation for Media QoS

- Two issues to support Media QoS for UMA in heterogeneous environments;
  - Adaptability
    - by exploiting content adaptation
  - Reliability
    - by exploiting content protection
- Contents are adapted into various qualities and modalities
  - by using modality conversion
    - e.g., video-to-image conversion
  - by using content transcoding (including transrating)
    - This allows the multimedia content delivery to adapt to the wide diversity of client device capabilities in communication, processing, storage, and display.
    - either for scalable video or non-scalable video
- Content Adaptation can be provided in two basic ways
  - by storing, managing, selecting, and delivering different versions of the media objects (images, video, audio, graphics and text) that comprise the multimedia presentations
  - by manipulating the media objects on-the-fly.

Content Distribution Model



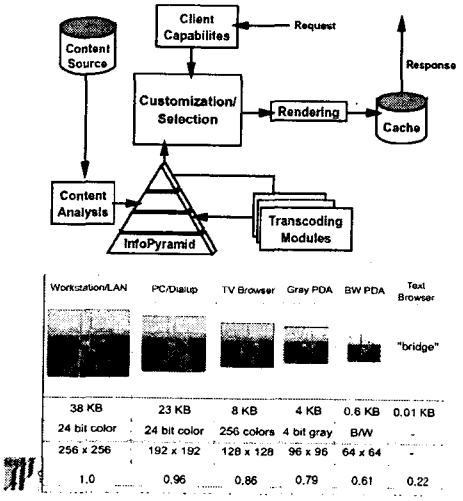
Key Issues on Content Adaptation



- Two basic questions in the decision engine;
  - Which is the modality of output content?
  - What is the quality of output content?

Related Works

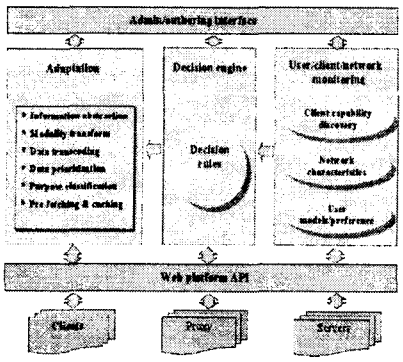
- InfoPyramid project (**IBM**)



- Content adaptation for pervasive devices such as (PDAs), mobile phones...
- Two key components:
  - A data representation scheme (InfoPyramid): provides a multi-modal and multi-resolution hierarchy for multimedia content
  - A content customizer: selects the variations of contents suitable to client capabilities, while providing the best representation to users

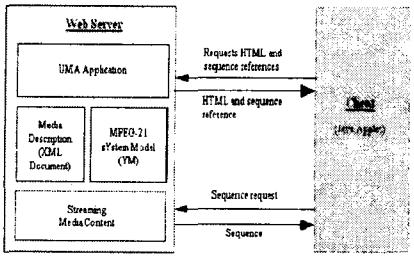
Related Works

- MPEG-21 DIA (Digital Item Adaptation)
  - Multimedia framework
- Cooltown (Hewlett-Packard)



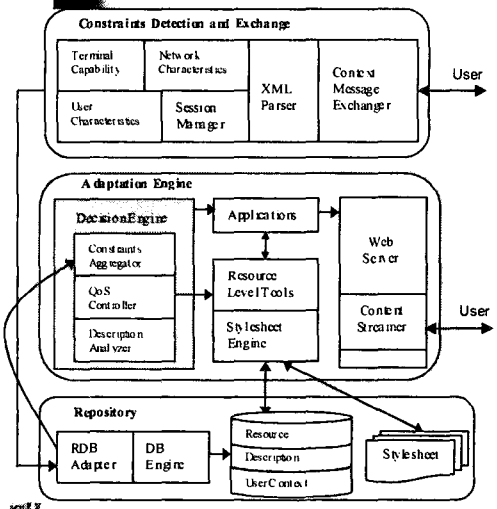
Source: HP

- Odyssey (Univ. of Michigan)
  - is to adjust the quality of accessed data to match available resources
- Midgard Media Lab (NTNU)
  - MPEG-21 compliant



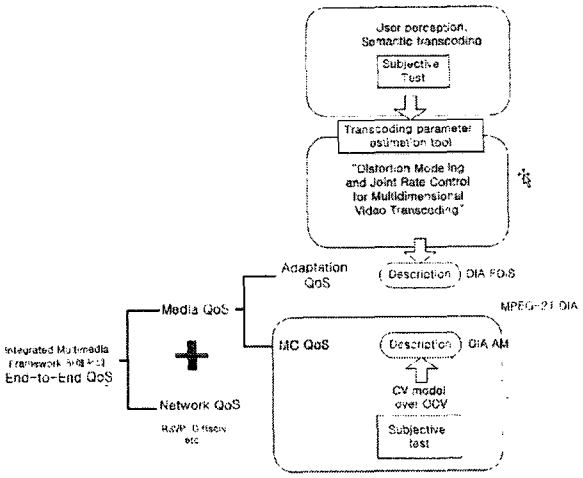
Source: NTNU

Content Adaptation System for Media QoS in ICU



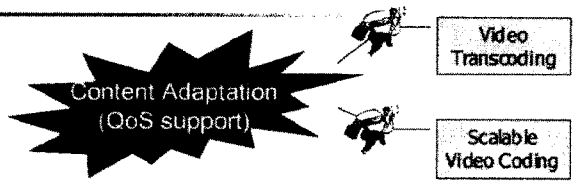
- ❑ Multimedia content adaptation system for the next multimedia framework
  - ❖ to various kinds of adaptation
  - ❖ MPEG-21 DIA compliant
- ❑ Constraints Detection and Exchange
  - ❖ Usage environment descriptions
    - ❖Network characteristics
    - ❖Terminal capability
    - ❖User characteristics
      - ❖ User preferences
    - ❖Session information
  - ❖ Context Message Exchange with User
- ❑ Adaptation Engine
  - ❖ 판단 엔진(Decision Engine)
  - ❖ 자원 적응변환 툴(Resource Adaptation Tool)
- ❑Repository
  - ❖ Storage for contents, descriptions, user context

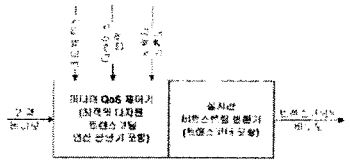
Research works



<E-E QoS 에서 미디어 QoS 관계 및 구체내용>

3. Video Transcoding for Media QoS





< Transcoding module >

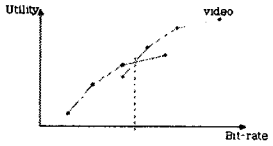
- Transcoding module tailors the quality of pre-encoded (stored) content according to various usage environments
  - Transrating, Format conversion (e.g, MPEG-2 → MPEG4)

Decision Engine for Video Transcoding

Targets:

- Two important questions to support QoS in Video Transcoding;
  - "When transcoding is to be happened?"
  - "How transcoding is to be?"
- Automatic decision for the best quality
  - Optimal Rate Control for Video Transcoding
  - Optimal Rate Control for Scalable Video Coding

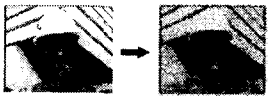
Quality Modeling for Decision Engine for Video Transcoding



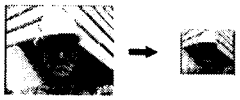
- Objective quality measurement
  - PSNR
  - MSE (mean square error)
- Objective measurement of subjective quality
  - The more sophisticated utility metrics that consider human vision systems capturing perception
  - Weighted PSNR, objective MOS
- Subjective quality measurement
  - User-satisfaction containing the semantic meaning of content
  - MOS (mean opinion score)
- Important issues;
  - Quality modeling through the analytical approaches,
  - Bridging the gap between the objective and subjective measures,
    - Semantic way
    - Perception way
  - Aggregation of multi-dimensional qualities.

Scaling Operations for Video Transcoding

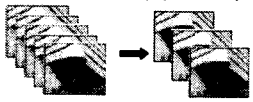
Spatial Quality(SNR) Scaling



Spatial Resolution Scaling



Temporal Resolution Scaling



30f/s ➡ 10f/s

- Why multidimensional transcoding?
  - ❑ Nowadays, at least 30~60 times or more rate adaptability is required.
    - ❖ Requantization (SNR dimension) operation provides limited reduction ratio.
  - ❑ To provide the best QoS for the given resource
  - ❑ QoS Tradeoffs

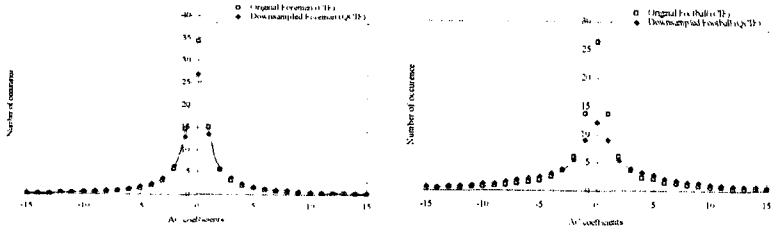
Joint Rate Control for Multidimensional Video Transcoding

- "Main objectives;
  - To determine the optimal combination of transcoding operations,
  - To jointly control the spatial, SNR, and temporal operations ,
  - To exploit spatio-Temporal-SNR tradeoffs.
- Approach
  - Rate control by using the multidimensional rate-distortion modeling
    - to estimate the target bitrate and distortion before performing the actual transcoding process
  - Rate control by considering inter-operation dependency
    - by *statistical analysis of video source*
- Previous works
  - Previous works focused on either bit allocation under a constant frame rate or frame rate control at a fixed spatial resolution
    - Entire R-D optimization for transcoding and the decision considering spatio-SNR-temporal tradeoffs are starting to be discussed.

Inter-operation Dependency

In ICIP2004

- Rate-Distortion Model
  - Statistical analysis of source



- Inter-operation dependency
  - Laplacian source distribution is affected by the amount of inter-operation dependency : $p(x) = \frac{\alpha'}{2} e^{-\alpha'|x|}$
  - Parameter alpha value of model for SNR is affected by TS and SS operations

Joint Rate Control Algorithm

In ICIP2004

- Joint Rate Control Algorithm for Hybrid Video Transcoding

Segment-level rate control algorithm for hybrid transcoding

Step 0: **Calculate the target bit rate** for a segment;

Step 1: **Estimate each bitrate** for all the possible combinations of transcoding operations which consist of the frame rate, the scaling factor of resizing, and the amount of average requantization for the segment;

Step 2: **Select the probable operation sets** from all the operation combinations in Step 1, which can generate the target rate;

Step 3: **Estimate the total distortion** for the selected operation sets in Step 2 by using the distortion measure;

Step 4: **Select an optimal operation set** which minimizes the total distortion by solving

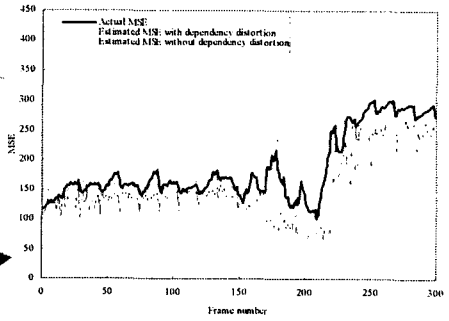
$$R_{th} < R_e < R_{\max}$$

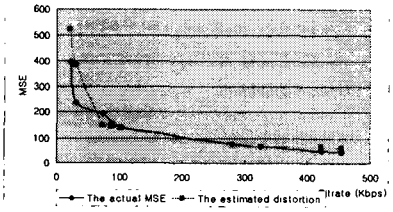
Video Transcoding

Experimental Result

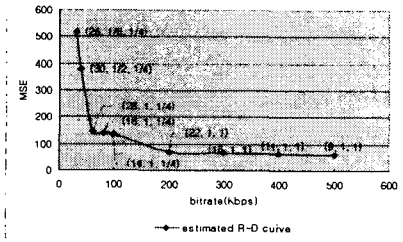
❑ Comparison of the estimated distortion and the actually measured MSE which are generated from CIF Foreman.

❑ Combination of transcoding operation (RQ, TS, SS) for Foreman sequence
- ❑ (18, 1, 1/4) for the 1st scene (frame 1 ~frame 171)
- ❑ (26, 1, 1/4) for the 2nd scene (frame 171 ~220)
- ❑ (25, 1, 1/4) for the 3rd scene (frame 221 ~300)





<Foreman Test Sequence>



Result of joint control for hybrid transcoding of CIF Foreman originally coded at 510Kbps.

Video Transcoding

Experimental Result

❑ *Transcoding operations used are as follows:*
- ➢ requantization parameter RQ={1,..,31}
- ➢ SS={1, 1/4}, i.e., {CIF,QCIF}
- ➢ TS={1, 2/3, 1/2, 1/3, 1/5}, i.e., {30f/s, 20f/s, 15 f/s, 10f/s, 5f/s}.

❑Container Ship has low motion ➔neighboring frames are very similar.

➔ Temporal downsampling causes a small amount of distortion

➔ The best operation set (29, 1/5, 1) based on the estimated MSE is a reasonable choice.

❑ On the contrary, Football has high motion.

➔A little amount of temporal downsampling is better.

➔Our result (14, 1, 1/4) matches well this estimation

Table 3. Estimated operations which can generate the rate 300Kbps from CIF Container Ship originally coded at 262Kbps

Average Requantization Parameter (RQ)	Frame Rate (TS)	Resizing Parameter (SS)	Estimated MSE
12	1/5	1/4	286.58
15	1/3	1/4	294.11
17	1/2	1/4	292.18
19	2/3	1/4	289.03
22	1	1/4	277.09
29	**1/5**	**1**	**165.50**

Best choice

Table 4. Estimated operations which can generate the target rate 300Kbps from CIF Football originally coded at 946Kbps

Average Requantization Parameter (RQ)	Frame Rate (TS)	Resizing Parameter (SS)	Estimated MSE
5	1/5	1/4	2920.88
7	1/3	1/4	2468.30
9	1/2	1/4	1909.77
11	2/3	1/4	1914.95
14	**1**	**1/4**	**1758.95**
14	1/2	1	2857.05
21	1/3	1	2410.32
27	1/2	1	1838.42

Best choice

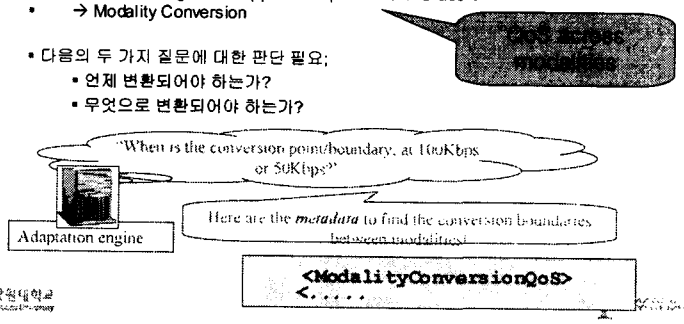
Advanced Video Transcoding

- Semantic Transcoding
  - aims to get the best "human experience" (quality of experience) in multimedia consumption
    - *not for just "universal access"*
  - Semantic connection between Human and Media
    - Quality (utility, fidelity) measure, modeling
      - Source rate control for best quality of experience
      - QoS trade-offs

4. Modality Conversion for Media QoS

Modality Conversion

- Modality
  - images, video, audio text, graphics, etc.
  - MPEG-1,2,4, H.263, JPEG...
- Motivation
  - The content scaling method has been used to support the quality of the content
  - The quality of the content, however, may be destroyed significantly
  - The content scaling can't support acceptable QoS to users
  - → Modality Conversion
  - 다음의 두 가지 질문에 대한 판단 필요;
    - 언제 변환되어야 하는가?
    - 무엇으로 변환되어야 하는가?

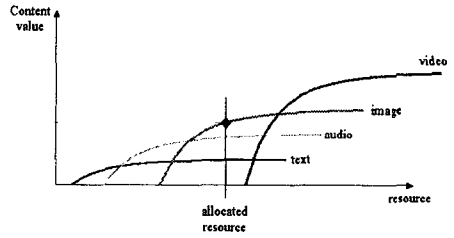


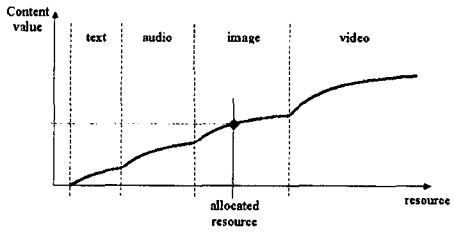
Four Factors in Modality Conversion

- Modality capability
  - which is the support for user's consumption of certain modalities.
  - This factor can be determined from the characteristics of terminal (e.g. text-only pager), user (e.g. blind user), and even surrounding environment (e.g. a too noisy place).
- User preference
  - shows user's levels of interest to different modalities.
- Resource constraints
  - for example: the terminal can support video modality but at some point the connection bitrate is not enough to play the video content online.
- Semantics of the content
  - for instance, between an interview video and a ballet video, the provider would be more willing to convert the former to a stream of text

Determining the Conversion Boundaries

- Concept of Overlapped Content Value (OCV) model
  - The aggregation of utilities in each QoS dimension and modality



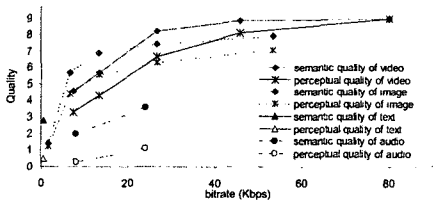


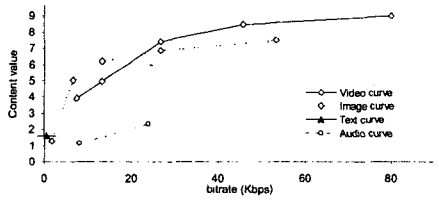
❖The intersection points are the conversion boundaries

❖Each modality curve has a scale factor showing the modality's importance

Content Value Modeling for Modality Conversion

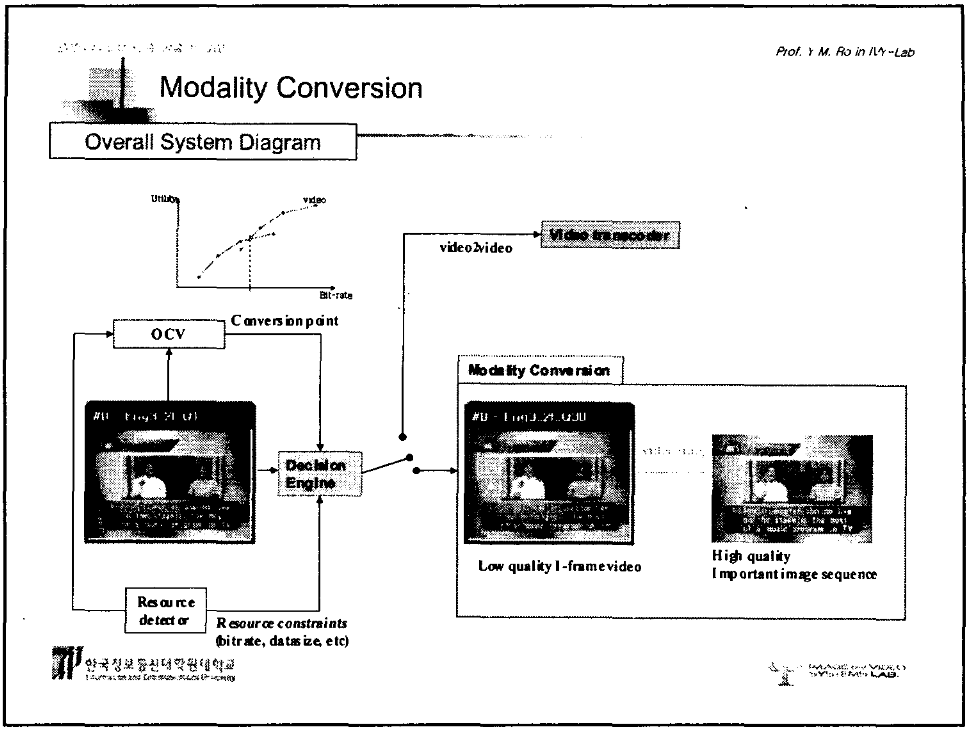
- Subjective quality measurement
  - the perceptual quality
    - e.g. lower frame rate
  - the semantic quality
    - understanding

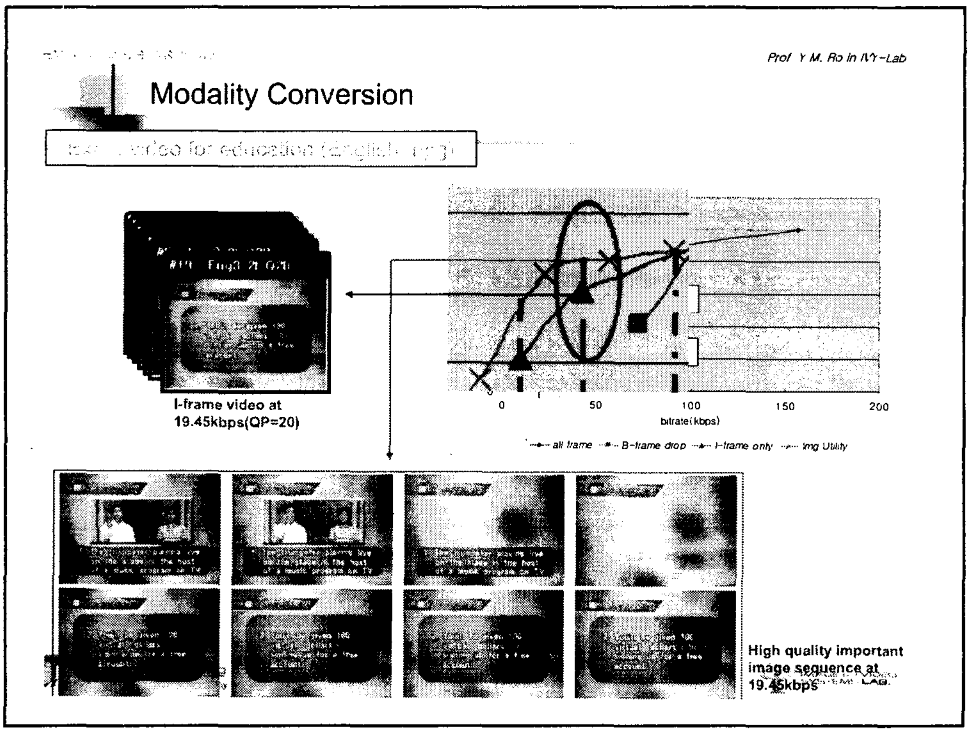




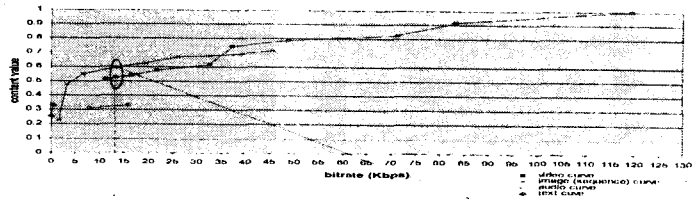
Content value

$V = s \times PQ + (1-s) \times SQ$


Prof. Y. M. Ro in IVY-Lab
Modality Conversion
Overall System Diagram
Utility
video
Bit-rate
OCV
Conversion point
video2video
Video transcoder
Modality Conversion
Decision Engine
Low quality I-frame video
High quality Important image sequence
Resource detector
Resource constraints (bitrate, datasize, etc)

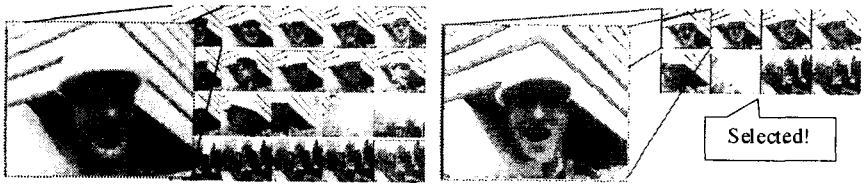

Prof. Y M. Ro in IVY-Lab
Modality Conversion
I-frame video at 19.45kbps(QP=20)
0
50
100
150
200
bitrate(kbps)
all frame
B-frame drop
I-frame only
img Utility
High quality important image sequence at 19.45kbps

Modality Conversion



Video stream having 20 frames of low quality

Image sequence having 8 images of original quality



Prof. Y. M. Ro in IVY-Lab

Modality Conversion

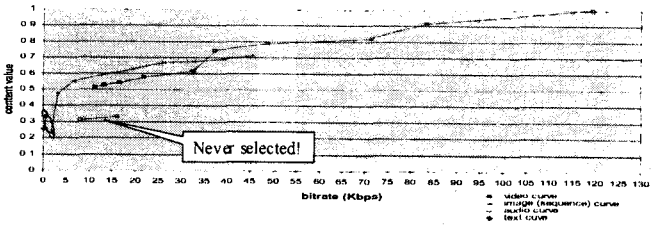


IMAGE VERSION

hat, a red shirt and a grey coat. There is house behind him. He seems to be very enthusiastic and happy. He is talking about something very important, maybe his work. He raises his hand to introduce. And he turns to the left. Our camera is paning right. Oh there is a big yellow crane! Here is certainly the construction site. They are building a house. There are many bricks

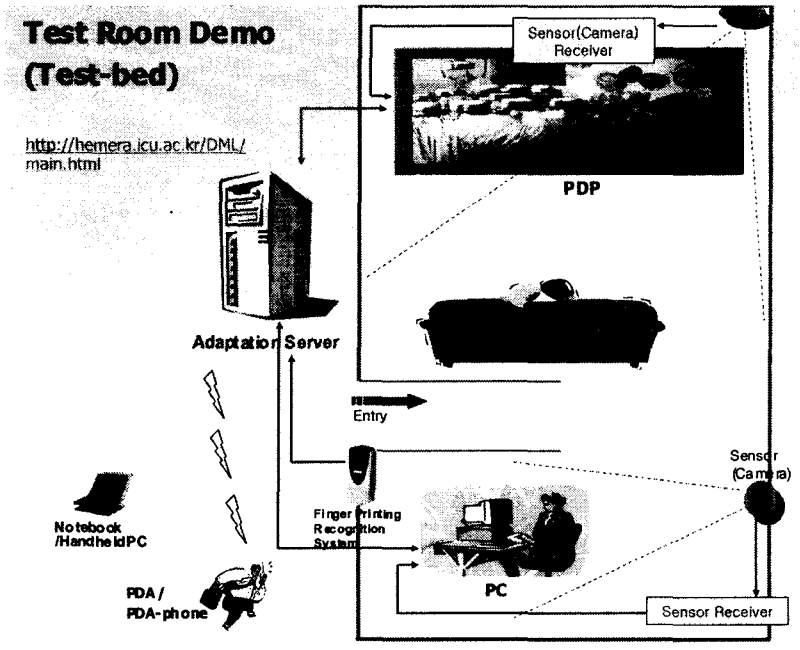
TEXT VERSION

Selected!

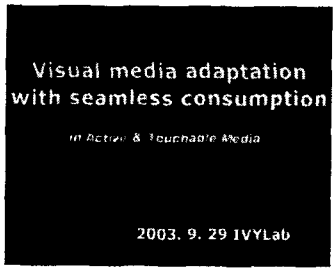
5. Demos for Content Adaptation

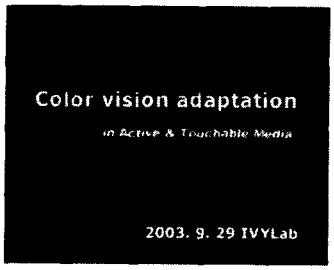
Test Room Demo (Test-bed)

http://hemera.icu.ac.kr/DML/main.html



Demo Video





http://hemera.icu.ac.kr/DML/main.html