CONTENT-ADAPTIVE UTILITY-BASED VIDEO ADAPTATION

Jae-Gon Kim, Yong Wang**, and Shih-Fu Chang***

*Electronics and Telecommunications Research Institute
Daejeon, Korea

**Department of Electrical Engineering
Columbia University, New York, USA

ABSTRACT

Many techniques exist for adapting videos to satisfy heterogeneous resource conditions or user preferences. However, selections of appropriate adaptations among various choices are often ad hoc. To provide a systematic solution, we present a general conceptual framework to model video entity, adaptation, resource, utility, and relations among them. The framework extends the conventional rate-distortion model in terms of flexibility and generality. It allows for formulation of various adaptation problems as resource-constrained utility maximization. It also facilitates new approaches to predicting critical information about resource-utility relations. We apply the framework to a practical case in dynamic bit rate adaptation. Furthermore, we present a description tool, which has been accepted as a part of the MPEG-21 Digital Item Adaptation (DIA), to support utility-based adaptation.

1. INTRODUCTION

Universal Media Access (UMA) represents an important technology area in the post-PC era. In the UMA environment, media content adaptation in response to variations of environment resources and user preferences is considered as a core technology.

Many adaptation methods exist for adjusting the bit rate of compressed video streams [1,2,4,8,9]. For example, requantization of transform coefficients, frame dropping (FD), DCT coefficients dropping (CD), and resolution reduction are commonly used. However, most existing works concentrate on optimization of pre-selected adaptations, rather than systematic solutions in choosing the optimal adaptations. In [2,8,9], the rate-distortion (R-D) characteristics were used to find optimal combinations of the frame rate and spatial quality of video. But the approaches do not seem readily extensible to different types of resources (other than the bit rate) and adaptations (other than quantization and frame skipping). Also, real-time generation and description of the R-D information was not addressed.

In this paper, we present a utility-based framework [3] that extends the conventional R-D model, by allowing flexible considerations of diverse types of adaptations, resources, and utilities. The resource-utility relationships represent analogous but broader concepts than the R-D relationships. Specifically, the framework models three important parameters generally represented in multidimensional spaces − *adaptation, resource,* and *utility (ARU)*, and relations among them (Figure 1).

Key issues arise in realizing the framework in practice: (1) identifying video entities and adaptations that are applicable to entities; (2) modeling resource and utility associated with each adaptation; and (3) selecting optimal adaptations at given constraints in resources and/or user preferences. To demonstrate the computational solutions to these issues, we present as a case study a hybrid transcoding combining FD and CD (referred as *FD-CD*) for dynamic bit rate adaptation of MPEG-4 videos, which is important in practical wireless mobile applications. The FD-CD adaptations have the benefits of implementation simplicity and computational efficiency. Both adaptations can be efficiently implemented in the compressed domain without full decoding of the compressed streams.

In addition, we present a representation scheme, called *utility function (UF)*, for describing the relations among ARU parameters.

The independent variables of UF include the sample points of the region of interest in the resource space, and permissible adaptations meeting resource constraints corresponding to each point. The value of the UF indicates the video utility after the adaptation is applied. Such UF information is very useful for selecting the optimal adaptation in a mobile adaptation engine located in a three-tier architecture depicted in Figure 2.

Section 2 describes an overview of the utility-based framework. Realization of the framework using the FD-CD adaptations is presented in Section 3. In Sections 4 and 5, we describe issues and solutions in predicting and describing the utility function. Current prototype status and conclusions are presented in Sections 6 and 7, respectively.

2. UTILITY-BASED ADAPTATION FRAMEWORK

2.1 Utility-Based Framework

Video adaptation problems involve identification of video content entity and the spaces of ARU. We use the term "space" in a loose sense here to indicate the multiple dimensionalities involved in each concept. For example, given a selected entity (e.g., a compressed video segment) there exist various adaptation methods such as re-encoding, requantization, FD, CD, temporal condensation, and resolution reduction etc. Resources are support from terminals or networks like bandwidth, computation capability, power, and display size, etc. Utility is the quality of video after undergoing adaptations. Utility can be measured in an objective or subjective manner. Many adaptation problems can be formulated as resource-constrained utility maximization based on the utility-based framework.

The ARU spaces are shown in Figure 1. The adaptation space is a conceptual space representing all possible adaptations. If we assume the adaptation method of FD and CD only, there are two dimensions: each dimension represents specific adaptation operations of FD and CD, respectively. Figure 1 also shows the associated mapping relations among the spaces. Given a video segment, a particular adaptation operation of FD-CD corresponds to a point in the adaptation space is associated with specific resources and utility values, which are represented by corresponding points in the resource space and the utility space respectively. Note that the mapping among the points in ARU spaces are often multiple-to-one since different adaptation operations may give the same resource values or induce the same utility value. The interesting point of the adaptation problem lies in choosing the optimal adaptation operation among multiple choices.

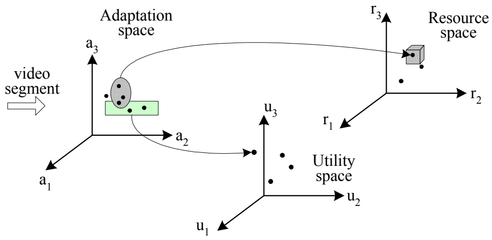


Figure 1. A general framework for modeling Adaptation-Resource-Utility (ARU) spaces in video adaptation.

2.2 A Three-Tier Architecture Utilizing the Framework

A three-tier (server-proxy-client) adaptation architecture, depicted in Figure 2, is considered as a promising solution to meet diverse resources and preferences in the UMA framework. Instead of re-generating a large number of videos with the same content but with different formats or qualities at the server, an adaptation engine is deployed in the proxy to dynamically adapt the video. Information about ARU relations is very useful for the adaptation engine to select the optimal adaptation operation. For live videos, generation of such utility information needs to be done in the real time (instead of offline). We will briefly describe a content-based real-time prediction method [5] in Section 4.

3. CASE STUDY: BIT RATE ADAPTATION

Let us illustrate the formulation and representation of the utility framework by considering a practical case of adapting MPEG-4 video to dynamic bit rates. The FD and CD are efficient and effective ways of rate adaptation with low computational complexity. The adaptations just truncate parts of the compressed bitstream and thus can be implemented to manipulate video frames and DCT coefficients directly in the compressed domain.

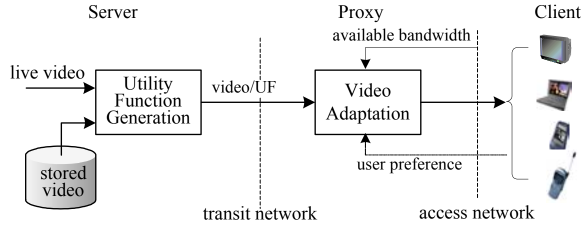


Figure 2. Three-tier adaptation architecture using the utility-based framework

3.1 Frame Dropping

FD is a common method of temporal adaptation that adjusts the frame rate by skipping frames in a video segment. We consider simple FD operations that allow only dropping the unreferenced frames within each group of pictures (GOP). If an anchor frame is dropped, subsequent frames may need to be re-encoded due to the concern of error propagation [1]. FD can meet only a coarse level of the target bit rate since the smallest unit of data that can be removed is an entire frame. For instance, in the case of GOP (N = 15, M = 3) (N: size of GOP, M: distance between two successive anchor frames), we have a set of discrete adaptation operations in the FD dimension: "no dropping", "one B frame dropping in each sub-GOP", "all B frames dropping", and "all B and P frames dropping". More permissible points can be added if we allow non-uniform dropping of frames in each GOP.

3.2 DCT Coefficient Dropping

We adopt a CD method similar to the previous work of dynamic rate shaping [4]. CD adjusts the spatial quality of each frame by truncating a subset of DCT run-length codes at the end of each block. Lagrangian optimization [4] was used to minimize the overall distortion by optimally allocating the number of DCT coefficients to be dropped in different blocks within a frame. Given a video segment and a target bit rate, we first employ uniform rate reduction across different frames. Then, within a single frame, we employ the above non-uniform dropping allocation among blocks. Unlike FD, CD provides the ability to meet the available bandwidth with a finer granularity by adjusting the amount of dropped coefficients. We can define the level of CD by specifying the percentage of rate reduction to be achieved. For example, "CD = 10%" represents an adaptation operation that reduces 10% of the bit rate in each frame by CD.

3.3 Combination of FD and CD

For significant bit rate reduction, FD or CD alone is not sufficient to accommodate the target rate. Moreover, only a few values of bit rates can be achieved by FD while finer rate adaptation is possible using CD. Therefore, the FD-CD combination enables us to expand the dynamic range of achievable bit rate while meeting the target rate at a finer level. Moreover, the FD-CD adaptation provides the freedom in balancing the trade-offs between spatial and temporal quality.

4. UTILITY FUNCTION REPRESENTATION AND GENERATION

We need a scheme to represent the relations among the permitted adaptation operations corresponding to the points in the adaptation space and the associated utility and resource values. If we consider only a single measure of resource (e.g., bit rate) and utility (e.g. PSNR) in the case of FD-CD, the similar scheme to the conventional R-D curve can be used. We call such representation *utility function*. An example of UF is shown in Figure 3. Different adaptation points (representing specific adaptation operations of FD-CD) are plotted on the resource–utility plane. In Figure 3, the adaptation points with the same FD level are grouped with the same curve. Different points on the same curve represent different amounts of rate reduction through CD (e.g., CD: 0 ~ 50%).

For stored videos, the UF can be generated by exhaustive off-line simulation since the computation time may not be a serious issue. But, for live applications, a real-time solution is needed in generating such UF information at the server or the proxy. Some prior works attempt to use analytical approaches in modeling the UF information [2,8,9] with certain levels of success. An alternative approach is to exploit the strong correlations between content characteristics and UF shapes and thereby develop content-based prediction methods. In [5], we have developed a content-based statistical approach that combines content feature extraction, statistical clustering, and regression for predicting the UF of live videos in real-time.

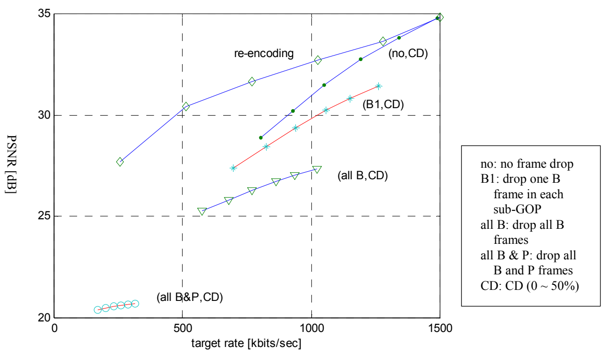


Figure 3. Utility function using FD-CD adaptations on the "coastguard" video (1.5 Mbps, GOP (N = 15, M = 3)).

5. DESCRIPTION OF UTILITY FUNCTION

We also developed a description scheme to represent the UF information in an interoperable way. The scheme shown in Figure 4 has been adopted as a part of *AdaptationQoS* descriptor in MPEG-21 DIA [7]. *AdaptationQoS* specifies a set of *Adaptation* descriptors, each of which describes an adaptation method (e.g., FD-CD or MPEG-4 scalability), and the elements of *Constraint*, *Utility*, and *UtilityFunction*. The *Constraint* and *Utility* respectively specify resource constraints and utilities measure in terms of name and unit. The *UtilityFunction* describes possible adaptation operators (e.g., a particular adaptation of FD-CD) and their associated utilities for a set of resource constraint points. The constraint points are discrete sample points of the region of interest in the resource space. The *LookUpTable* is used to describe additional information that may be needed in more elaborate scenarios in a form of table.

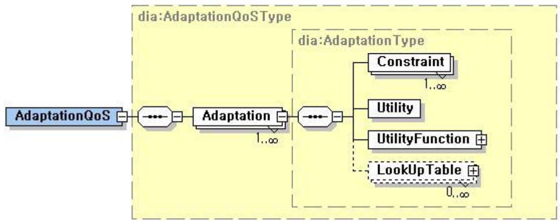


Figure 4. Schema diagram of *AdaptationQoS* descriptor.

Addressing ambiguity in specifying adaptation operations

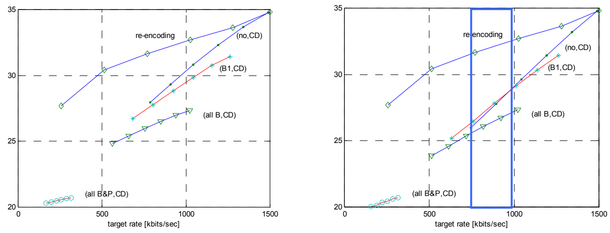
Adaptation operations may not be unambiguously defined in some adaptation method. For example, an operation of "CD = 10%" does not specify the exact set of coefficients to be dropped. Different implementations may choose different sets and result in slightly different utility values. On the other hand, for scalable compression formats such as JPEG-2000 and MPEG-4 FGS, part of the scalable stream can be truncated in a consistent way without ambiguity.

To address the ambiguity issue, we introduce the notion of utility ranking in our UF representation scheme. We assume that, in some applications, the relative rankings of utility values among different adaptation operations meeting the same resource are more useful than the absolute values of utility. The likelihood of getting consistent utility rankings from different implementations is higher than getting consistent absolute utility values. Such options of describing utility ranking allow the applicability of the *AdaptationQoS* to implementation-dependent adaptation methods.

Figure 5 shows the variations of UF resulting from different implementations of CD using the same experimental conditions shown in Figure 3. Specifically, we have implemented three different CD algorithms:

- Figure 3 – optimized CD that minimizes overall distortion by Langragean optimization in each block;
- Figure 5 (a) – the same optimized CD as above but optimization is done in each macroblock (MB);
- Figure 5 (b) – drop the same percentage of coefficients in each block without R-D optimization.

We observe that there are indeed noticeable variations of utility values among different implementations. However, in most of the bit rate range (except the boxed area in Figure 5 (b)), the utility rankings among competing adaptation operations at a given bit rate are consistent among different implementations. Even within the exception range, some operators may have consistent rankings (e.g., the operator of (all B, CD) has the worst utility regardless of implementations).



(a) (b)

Figure 5. Utility functions resulting from different implementations of CD (a) MB-level optimized (b) uniform dropping without R-D optimization.

6. EXPERIMENTAL RESULTS

We have developed a prototype of the proposed utility-based video adaptation system that uses FD-CD to adapt MPEG-4 video streams in response to dynamic bandwidth constraints. The prototype consists of a metadata parser, an adaptation engine, and a user interface. The parser extracts the UF descriptor associated with a video segment. The adaptation engine chooses the optimal adaptation operator according to the descriptor at the given target bit rate, and transcodes the incoming video clip. The user interface allows monitoring of the adapted video, the UF, the chosen operator, and the trace of the adapted bit rate to the time-varying target rate.

For live videos requiring the dynamic prediction of UF, our statistical predictor based on content characteristics [5] achieves the accuracy more than 80% on average in estimating the optimal combination of FD-CD over a wide range of bandwidth, which is regarded very promising. The utility-based adaptation also achieves promising performance gain (0.83 dB in PSNR and noticeable improvement on subjective quality) over the conventional system, which is based on a priori probability without being guided by UF, in primary testing of actual videos.

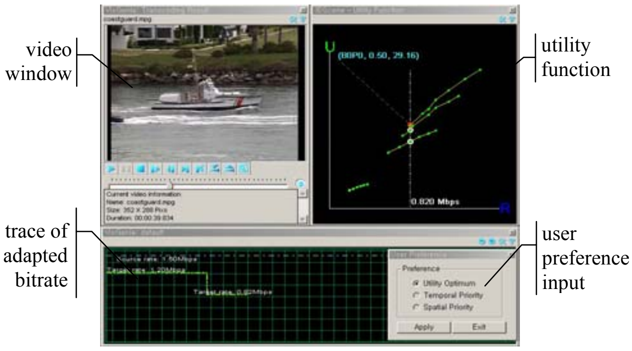


Figure 6. User interface of the prototype of utility-based adaptation.

7. CONCLUSIONS

This paper describes a general framework and a systematic methodology for video adaptation to meet diverse resource conditions and user preferences in UMA environments. The framework explicitly models the major concepts involved in adaptation processes – adaptation, resource, and utility. It resembles the conventional R-D model, but allows for greater flexibility in combining diverse types of adaptations, resources, and utility measures. We also present an efficient description scheme, called utility function, for representing critical utility-resource information needed in selecting the optimal adaptation operation meeting given resource constraints. Such representations facilitate a parametric form that can be accurately clustered and predicted by techniques we develop in [5]. We demonstrate the realization of the proposed methods in a practical case of MPEG-4 video bit rate adaptation. Our experiments show promising performance gain by using the utility-based adaptation. Finally, we have submitted to MPEG-21 DIA the description scheme, called AdaptationQoS, which has been accepted as a part of the Committee Draft.

ACKNOWLEDGEMENTS

This work has been partly supported by the Ministry of Information and Communication of Korea.

REFERENCES

[1] T. Shanableh and M. Ghanbari, “Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats,” *IEEE Trans. Multimedia*, vol. 2, pp. 101-110, June 2000.

[2] A. Vetro, Y. Wang, and H. Sun, “Rate-distortion optimized video coding considering frameskip,” in *Proc. IEEE ICIP*, Thessaloniki, Greece, Oct. 2001, pp. 534-537.

[3] S.-F. Chang, “Optimal Video Adaptation and Skimming Using a Utility-Based Framework,” in *Proc. IWDC-2002*, Capri Island, Italy, Sep. 2002.

[4] A. Eleftheriadis and D. Anastassiou, “Constrained and General Dynamic Rate Shaping of Compressed Digital Video,” in *Proc. IEEE ICIP*, Washington, DC, Oct. 1995, pp. 396-399.

[5] Y. Wang, J.-G. Kim, and S.-F. Chang, “Content-based utility function prediction for real-time video transcoding,” submitted to ICIP 2003.

[6] J.-G. Kim, Y. Wang, S.-F. Chang, K. Kang, J. Kim, “Description of utility function based optimum transcoding,” ISO/IEC JTC1/SC29/WG11 M8319, Fairfax, May 2002.

[7] *MPEG-21 Digital Item Adaptation CD,* ISO/IEC JTC 1/SC 29/WG 11/N5353, Dec. 2002.

[8] J.-J. Chen and H.-M. Hang, “Source model of transform video coder and its application— Part II: Variable frame rate coding,” *IEEE Trans. Circuits Syst. Video Techno.*., vol. 7, pp. 187-298, Apr. 1997.

[9] E.C. Reed and J.S. Lim, "Optimal multidimensional bit-rate control for video communications," *IEEE Trans Image Processing*, vol. 11, no. 8, pp. 873-885, Aug. 2002.