Optimal Video Adaptation for Resource Constrained Mobile Devices Based on Utility Theory

Özgür D. Önür[1,2], A. Aydın Alatan[1,2]
[1]Department of Electrical and Electronics Engineering, M.E.T.U.
[2]TÜBİTAK BİLTEN, Balgat 06531 Ankara TURKEY

ABSTRACT

The diversity of user terminals that can be used to access multimedia content necessitates tailoring of the content according to the computational capabilities of the terminals. Optimal video adaptation based on multimedia playback device characteristics is addressed by using novel approach. The adaptation is optimal in the sense that, the adapted video maximizes the user satisfaction. Utility Theory is used to formulate the satisfaction a user gets from watching a video. The proposed approach divides the user satisfaction into three independent components. The individual components are modeled by exponential curves and their weighted sum is used as the overall satisfaction or 'utility' function. The combination of trade-off weights that result in highest user satisfaction are obtained through experiments with different device characteristics. Experimental results indicate promising performance in terms of subjective quality.

1. INTRODUCTION

One of the major challenges, obscuring the path towards enjoying mobile multimedia, is delivering the multimedia content towards various mobile terminals across a wide range of networks [1]. This issue has attracted considerable attention in signal processing community and the concept of Universal Multimedia Access (UMA) has been devised to deal with this challenging problem [2]. The solution to UMA requires adjusting the resource requirements of video as it traverses along the interconnected networks from the satellite to the mobile terminal, since different networks and terminals will certainly have diverse capacities and characteristics. The process of modifying a given representation of a video into another representation, in order to change the amount of resources required to transmit, decode and playback the video is called as *video adaptation* [1-3].

In this paper, a novel method to determine an optimal video adaptation scheme, given the properties of an end terminal on which the video to be displayed, is proposed. *Utility Theory* [6] is utilized to construct models, which are fitted to results of subjective evaluation tests, formulating the "satisfaction" a user gets from watching a certain video clip. The optimal video adaptation is achieved by maximizing a '*utility function'* to determine the representation of the video that results in the highest user satisfaction.

2. RELATED WORK

As one of the earliest examples, a system to adapt multimedia web content to match the capabilities of a requesting device is introduced in [2] with an *InfoPyramid,* which creates and stores, multimodal and multi-resolution representations of the multimedia content. Using this representation, a "customizer" selects the representation of the content from the various available versions. Considering the diversity of the terminals that can be used to access multimedia content, an optimal representation for each of these terminals cannot be obtained from a predetermined set of representations by this method.

The first reference to *utility theory* in the context of video adaptation appears in [3]. In a more theoretical approach, only a conceptual framework to model adaptation, as well as resource, utility and the relationships in between are presented [4]. While objective measures, such as PSNR, coherence, temporal smoothness are used to measure utility [4], the optimal video adaptation problem is formulated as finding the adaptation operation that maximizes the utility of the adapted *entity* [4], given the original entity and resource constraints. However, the objective measures fail to model human satisfaction adequately. Hence, obtaining an acceptably accurate model a multitude of attributes need to be extracted from the video, and this significantly increases the complexity of the system.

Recently, a content-based utility function predictor is proposed in [5]. The system extracts compressed domain features in real time and uses content-based pattern classification and regression to obtain a prediction to the utility function. Nevertheless the utility value corresponding to a given adaptation of a video is presented as a function of the video bitrate, which contradicts the subjective nature of the utility concept.

3. UTILITY THEORY

The fundamental motive of utility theory is, to represent the satisfaction or *expected utility* of a resource, as a function of the amount of that resource [6]. There are two methods to obtain the utility function of a resource in utility theory. Both of these methods rely on subjective utilities provided by individual(s) representing the community for which the utilities need to be determined. While one of the approaches requires eliciting the utility values directly from the individual, by presenting the best and worst possible results and asking the individual to determine the relative satisfaction of *all* the remaining points of the utility function, the other approach assigns a specific shape (usually an exponential) to the utility function, such as [6]

$$U(x) = x^{c} \quad \text{or} \quad U(x) = (1 - e^{-x/c}) \tag{1}$$

where x stands for resources. The value of the parameter c should be estimated again by subjective tests, but this problem can be accomplished by using much less points.

In some cases, it might also be necessary to consider multiple objectives when trying to find the utility associated with an alternative. In other words, the total satisfaction might depend on more than one kind of resource. In such a case, if the satisfaction on any one of the objectives (also termed *sub-objectives*) is independent from the satisfaction from every other objective, the *additive utility function* [6] can be used to obtain the total satisfaction, as

$$\begin{bmatrix} Total \\ Satisfaction \end{bmatrix} = w_1\, U(obj_1) + w_2\, U(obj_2) + ... + w_n\, U(obj_n) \tag{2}$$

It is obvious that the satisfaction a user will get from viewing a video clip depends on more than one kind of resource (e.g. bit rate, frame rate etc.). Hence, if this satisfaction can be decomposed into sub-objectives, which are independent, the total user satisfaction can be modeled as the sum of these terms. The benefit of such decomposition is an easier determination of individual sub-objectives by using subjective tests, rather than trying to model the total satisfaction as a whole.

4. PROPOSED SYSTEM

The main aim of this paper is to accurately determine the "satisfaction" a user gets from watching a video clip on a resource limited device, as a function of video coding parameters and the terminal device properties. For a given terminal (in this work, only CPU and the screen size values are considered), the user satisfaction is evaluated, as the video coding parameters, i.e. bit rate, frame rate and spatial resolution are varied. Construction of a utility function for this problem requires conducting quite a large number of experiments even if the second method illustrated by (1) is used. Since the utility is a function of 5 different variables, expressing utility as a simple exponential function as in (1) is actually not feasible.

In this manuscript, a novel approach to obtain the utility function for the above problem is proposed. The problem is considered as a multiple objective utility formulation. The overall utility function is decomposed into 3 independent components such that the satisfaction associated with any one of the components is independent from every other component. These components are determined as: "crispness" utility of a video clip; "motion-smoothness" utility of a video clip and finally, the spatial resolution utility of a video clip

The reason of such decomposition is due to their perceptual independence. In other words, video frames with very low distortion might be displayed in a non-smooth manner in time or a motion smooth video can independently have a very low spatial resolution. Independent analysis allows the sub-objectives to be expressed as simple functions of the video coding parameters.

4.1. Crispness Utility

The most accurate measure of crispness of a video might be the number of encoded bits per pixel (bpp). In order to express the encoded bpp in terms of the coding parameters, the bit-rate needs to be normalized by both frame rate and spatial resolution. Hence, the first component of the overall utility function can be formulated as

$$\begin{aligned} Ucrisp(coded\ \ bits\ per\ \ pixel\ ) &= U_{crisp}\ (CBR/(CFR\ \ CSR)\ ) \\ &= U_{crisp}\ (CBR, CSR,\ \ CFR\ ) \end{aligned} \qquad (3)$$

where CBR stands for Coded Bit Rate, CSR stands for Coded Spatial Resolution, and CFR stands for Coded Frame Rate. It should be noted that all video coding parameters are referred as *coded* parameters, since video parameters can be viewed differently when video is rendered on a resource limited device. The phrase "coded" is used to emphasize that the values being used here, are the original encoding values of the parameters, forced at the encoder. Finally, it should be noted that crispness should be additionally related to CSR, since for a given display device with a given screen size, smaller images tend to be perceived more crisp compared to larger images [7].

The user satisfaction for crispness of a video, should increase substantially, as bpp value is increased. However, this increase is expected to reach to saturation after a certain value of the bpp. This saturation is due to the inability of the Human Visual System (HVS) to discern the difference in crispness of a picture, resulting from increasing the bpp value beyond a certain point [7]. In the light of the above observations and (1), it can be asserted that, the utility of crispness curve, should have an exponential form as expressed by the following formula

$$U_{crisp}(CBR, CSR, CFR) = 1 - e^{-c_1 \frac{CBR}{CFR\ CSR}}, \quad c_1\ \alpha\ CSR \qquad (4)$$

Note that value of c_1 affects the rate of decrease of the exponential in (4). Since, the bpp value, required to code a picture for a given crispness, can be assumed to be smaller for the pictures having higher spatial resolutions, c_1 has been included in the above formulation as a function, in order to account for this fact, being directly proportional with CSR.

4.2. Motion Smoothness Utility

The motion smoothness of a video clip can be simply characterized by the coded frame rate in an infinite resource device. The frame rate, at which the observed frame rate in a resource limited device, deviates from the original coded frame rate also depends on the encoded video bitrate. This is expected, since decoding a high bitrate video requires significant computational resources, and after a certain bitrate is exceeded, the CPU will not be able to decode the video in real time. In light of the above discussion, it can be stated that the motion smoothness of a video being observed on a user terminal, should depend on the frame rate at which the video was originally coded, the bit rate of the video, and CPU of the end terminal. Thus, the second component of the utility function is determined as

$$U_{smooth}(CFR, CBR, CPU)$$

Intuitively, increasing CFR, this second component of the utility function should also increase up to a point as an exponential expression in a similar form to the crispness utility in (4). The point at which the utility of motion smoothness starts decreasing, due to resource limitations, should depend on the CBR of the video, as stated earlier. Hence, the smoothness utility can be modeled as follows: a function FR (CBR), determines the exact location of the "turning point"; i.e. the frame rate FR (CBR) at which the motion smoothness starts decreasing, while increasing CFR, for a given bitrate. The formulation of the dependence of motion smoothness utility on the CPU of the terminal device, is also simplified as assuming only two clock-speeds, which are *CPU Low* and C*PU High*. Based on these reasonings, the following utility function is proposed:

$$U_{smooth}(CFR, CBR, CPU) = \left\{ \begin{array}{ll} 1 - e^{-a_0\ CFR} & CFR \le FR_L(CBR) \\ a_1\, e^{-c_2(CFR - FR_L(CBR))} & CFR > FR_L(CBR) \end{array} \right\} CPU\ Low$$

$$U_{smooth}(CFR, CBR, CPU) = \left\{ \begin{array}{ll} 1 - e^{-a_0\ CFR} & CFR \le FR_H(CBR) \\ a_1 + 1 - e^{c_3(CFR - FR_H(CBR))} & CFR > FR_H(CBR) \end{array} \right\} CPU\ High \qquad (5)$$

$$FR\ \ \alpha\ \ \frac{1}{CBR} \qquad\qquad a_1 = 1 - e^{-a_0\, FR}$$

For the frame rates, where the utility is increasing (up to the limit defined by FR), the utilities of the high CPU and the low CPU cases are assumed to be the same. This assumption is reasonable, since the observed and the coded frame rates are same up to that point, and a particular frame rate gives the same utility across all platforms unless distorted by the resource constraints. Note that, different FR's (FR_L, FR_H) are used for high and low CPU cases.

The term a_0 in (5) is a constant to be determined based on the results of the subjective tests and intuitively, FR should be inversely proportional to CBR, as already explained. On the other hand, a severe degradation in motion smoothness utility is expected, as soon as CBR value increases beyond the decoding capacity of a CPU. In order to account for this fact, two functions c_2 and c_3 are used in the above formulation. Notice that in both expressions of (5), for larger c_2 or c_3, the utility drops faster. Hence, selecting c_2 and c_3 in direct proportionality to CBR, the desired form for the utility curves can be obtained. The point a_1 in both expressions is the value of utility, at which the functions start decreasing at the frame rate FR.

4.3. Spatial Resolution Utility

Intuitively, the utility of the spatial resolution of a video clip should depend on two factors: Initially coded spatial resolution of the video and the *screen size* of the user terminal. One can easily agree that a video, being transmitted to a terminal whose screen size is smaller than the CSR of this video, can only be viewed partially, i.e. clipped before being displayed on that device. This will inevitably result in reduced user satisfaction and should be avoided, if possible. The final component of the utility function is prototyped as follows:

$$U_{size}(CSR, Screen\ Size)$$

The utility of the spatial resolution of a video clip is expected to increase in a similar fashion to (4) and (5), up to the point at which the spatial resolution becomes equal to the screen size of the terminal. After that point, the utility is expected to decline conforming to the following equation:

$$U_{size}(CSR, ScreenSize) = \begin{cases} 1 - e^{-a_{21} CSR} & CSR \le ScreenSize \\ c_4 e^{-a_{22}(CSR - ScreenSize)} & CSR > ScreenSize \end{cases} \tag{6}$$

$$c_4 = 1 - e^{-a_{21} ScreenSize} \qquad a_{21} \ \alpha \ \frac{1}{ScreenSize} \qquad a_{22} \ \alpha \ \frac{1}{ScreenSize}$$

The parameters a_{21} and a_{22} are both inversely proportional with the screen size of the terminal. Note that, larger a_{21} leads to a steeper increase, in the increasing portion of the utility function, and larger a_{22} means a steeper decrease in the declining portion of the utility. Since increase and decrease in utility is expected to change more abruptly in smaller screens, the inverse proportionality of a_{21} and a_{22} is reasonable.

4.4. Utility Function Generation

The satisfaction for each of these sub-objectives is assumed to be independent of the satisfaction on every other sub-objective. For example, the satisfaction a user gets from the motion smoothness of a video has no dependency on the crispness of the same video. Therefore, one can use the *additive utility function* [6] to determine the total satisfaction a user will get from viewing a certain video. Thus the resulting equation is as follows:

$$U = w_1 U_{crisp} + w_2 U_{smooth} + w_3 U_{size} \tag{7}$$

The weights, w_1, w_2 and w_3, associated with the terms of the utility function, are to be determined by using simulations.

4.5 Subjective Tests for Utility Function

At this step, the unknown parameters of the utility expressions are determined by a series of subjective evaluation experiments. These experiments are performed separately for each component. While an experiment on one of the components is being performed, the video coding parameters not affecting the utility of that component are kept constant. The experimental methodology and environment are chosen to be analogous to the cases in subjective video evaluation standards [8].

The evaluators are first shown the videos that are considered the best and the worst for the particular component of the utility function. The evaluators are shown videos, coded with different values of the parameter(s) that has an influence on the component of the utility being tested. Then, the subjects are asked to grade those samples, according to the satisfaction they get from viewing that video. The important point here is that they are asked to evaluate the videos, only according to the component (e.g. crispness) being tested. After subjective values are collected, the unknown parameters are all determined using the results of the above tests by least squares fitting. For maximizing obtained utility function, a well-known stochastic optimization technique, *simulated annealing,* is used [9].

5. SIMULATIONS

5.1 Utility Curves for Sub-objectives

The utility curves obtained for the individual sub-objectives from the subjective tests are presented in Figs. 1, 2 and 3. As shown in Fig. 1, the crispness utility tests are performed for 4 different spatial resolutions and the utility was obtained as a function of coded bits per pixel. In all cases, the exponential increase and the preceding saturation, which are predicted by the proposed model, can be observed.

In Figure 2, only the motion smoothness tests for the *CPU High* case are shown. These tests are performed for 5 different bitrates and the results show the expected decrease at different frame rates depending on CBR.

In Figure 3, the results of utility of video size experiments are presented. The experiments are performed for 3 devices having different screen sizes and the utility curve starts exponentially decreasing as soon as videos with larger size than the screen size are being displayed.

5.2 Utility Values for General Terminals

The final stage in obtaining the complete utility function is determining the values of the trade-off weights, w_1, w2, w3, used in (2) at which the utility function has the maximum value. A series of simulated annealing experiments are performed, for four different combinations of Screen Size and CPU values (400Mhz, 200Mhz) X (352x288, 176x144), where the capital X denotes Cartesian product. A weight space is defined as W^3 where $\{w \in W \rightarrow w \in [0,1]\}$ and is discretized to steps of 0.1. An additional condition w1+w2+w3 = 1 is used to constrain the weight space. The experiments are performed by inserting all the possible combinations of weights w1,w2,w3 as defined by the above weight space into the formulation and observing the value of the utility function for each of the possible combinations. It is seen that the value of the weights that maximize the utility function do not change for different user terminals. Although the terminals, on which the experiments, are performed do not span the entire range of terminals, the results are expected to be approximately the same. When the weights are obtained, the utility function is uniquely determined. Hence, the optimal values of the video encoding parameters can be found for any terminal device. The system requires only the CPU and the screen size of a given terminal device to compute the values of

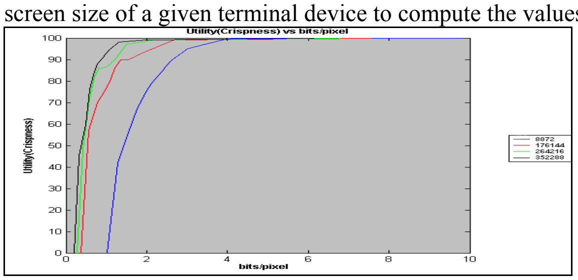

Figure 1: Utility of Crispness vs. bits/pixel

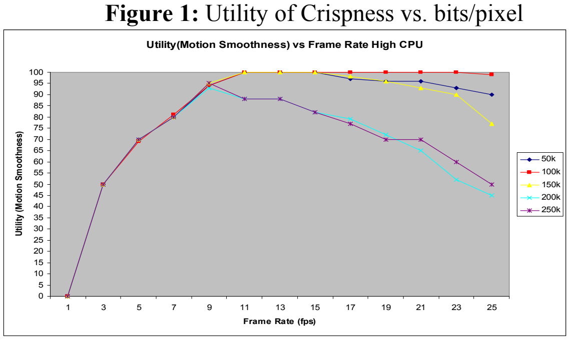

Figure 2: Utility of Motion Smoothness vs. Frame Rate

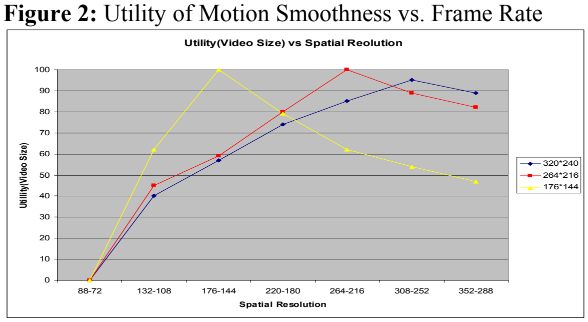

Figure 3: Utility of Spatial Resolution Size vs. Screen Size

the video coding parameters for the most satisfying viewing. It should be emphasized that by performing these experiments for only a limited set of user terminals, the necessary formulation to determine the optimal video coding parameters for a large variety of user terminals are obtained.

Figure 4 – A typical device with CPU=400MHz and screen resolution 320x240; optimal encoding parameters: Spatial Resolution 321x241, Frame Rate 12 fps, and Bit rate 182Kbits/s

Figure 4 shows a typical device on which the proposed method is applied. Some other typical results for different terminals are also presented in Table 1.

INPUT		OUTPUT		
CPU	Scr. Size	Bitrate	Fr. rate	Sptl. Res.
400 Mhz	320*240	182Kbps	12	321*341
400 Mhz	176*144	129Kbps	18	176*144
200 Mhz	320*240	80Kbps	10	320*240
200 Mhz	176*144	82Kbps	10	183*137

Table 1. Input and output parameters after utility-based optimization

6. CONCLUSIONS

The main contribution of this paper is the decomposition of the satisfaction a user gets from watching a video into three conceptually independent components, as the satisfaction resulting from the crispness of a video, the satisfaction resulting from the motion smoothness of a video and the satisfaction resulting from the spatial resolution of a video. It has been observed that such decomposition enables more accurate subjective evaluation of the user satisfaction. This in turn makes possible, precise modeling of the user satisfaction in terms of the video coding parameters. The proposed system is tested on a typical device and resulted with promising performance for user viewing satisfaction on such a device.

7. REFERENCES

[1] Y. Neuvo , J. Yrjanainen, "Wireless Meets Multimedia – New Products and Services," IEEE ICIP 2002.

[2] R. Mohan, J. R. Smith, C.-S. Li, "Adapting Multimedia Internet Content for Universal Access ," IEEE Trans. on Multimedia, vol. 1, no. 1, March 1999.

[3] P. Bocheck, Y. Nakajima and S.-F. Chang, "Realtime Estimation of Subjective Utility Functions for MPEG-4 Video Objects," Proc. of IEEE Packet Video Workshop (PV'99), New York, USA, April, 1999.

[4] S. F. Chang, "Optimal Video Adaptation and Skimming Using a Utility-Based Framework," Tyrrhenian International Workshop on Digital Communications (IWDC-2002) Capri Island, Italy September 2002.

[5] Y. Wang, J.-G. Kim, S.-F. Chang, "Content-Based Utility Function Prediction For Real-Time MPEG-4 Video Transcoding," IEEE ICIP 2003

[6] A. L. Golub, "Decision Analysis: An Integrated Approach," John Wiley and Sons. Inc., 1997

[7] Recommendation ITU-R BT.500-11 "Methodology for the Subjective Assessment of the Quality of TV Pictures," 2002

[8] K. Boff, L. Kaufman, & J. Thomas (Ed.), Handbook of Perception and Human Performance New York: Wiley, 1990.

[9] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, "Optimization by Simulated Annealing," Science, vol. 220, May 1983.