Video Adaptation for Transmission Channels by Utility Modeling

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 satisfaction a user gets from watching a video in a resource limited device, can be formulated by Utility Theory. The resulting video adaptation is optimal in the sense that the adapted video maximizes the user satisfaction, which is modeled through subjective tests comprising of 3 independent utility components : *crispness*, *motion smoothness* and *content visibility*. These components are maximized in terms of coding parameters by obtaining a Pareto optimal set. In this manuscript, inclusion of transmission channel capacity into the subjective utility model of user satisfaction is addressed. It is proposed that using the maximum channel capacity as a restriction metric, certain members of the Pareto optimal solution set can be eliminated such that the remaining members are suitable for transmission through the given channel. Once the reduced solution set is obtained, an additional figure of merit can be used to pick a single solution from this set, depending on the application scenario.**

I. INTRODUCTION

The process of modifying a given representation of a video into another representation, in order to change the amount of resources required for transmitting, decoding and displaying video is defined as *video adaptation* [1]. The first reference to *utility theory* in the context of video adaptation appears in [2]. In a more theoretical approach, only a conceptual framework that models adaptation, as well as resource, utility and the relationships in between, are presented [3]. A content-based utility function predictor is also proposed [4], in which 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. However, the utility value, corresponding to a given adaptation of a video, is presented as a function of the video bit-rate [4], which contradicts the subjective nature of the utility concept.

In [10], a novel method to determine an optimal video adaptation scheme, given the properties of an end-terminal, on which the video is to be displayed, is proposed. *Utility Theory* [5] is utilized to model a strictly subjective quantity, *satisfaction*, a user will get from watching a certain video clip. The satisfaction is formulated as comprising 3 independent utilities, each depending on certain video coding parameters.

In this manuscript, the effect of transmission channel capacity on the previously proposed subjective models [10] is addressed. The incorporation of the channel capacity effect to the models in [10] results in a complete formulation of the overall user experience in a multimedia delivery scenario from the content server to the mobile user terminal, as depicted in Figure 1.

II. PROPOSED ADAPTATION SYSTEM

The main aim of this paper is quantitatively determining the "satisfaction" a user gets from watching a video clip on a resource limited device, as a function of video coding parameters, the terminal device properties and also the capacity of the communication channel. Initially, the subjective satisfaction of the user is modeled ignoring channel capacity [10]. Subsequently, the effects of finite channel capacity are incorporated into the proposed model.

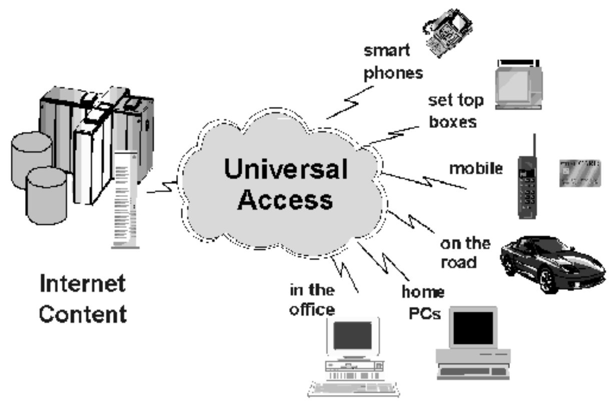

Figure 1. Delivery of the multimedia content from the media server to mobile terminal

A novel approach to obtain the utility function for the problem above, 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 these terms can be considered as independent from every other component. These terms are determined as:

- "Crispness" utility of a video clip,
- "Motion-smoothness" utility of a video clip,
- "Content visibility" utility of a video clip.

The reason for such decomposition is due to the perceptual independence of the proposed sub-objectives. 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. When the above decomposition is performed, the sub-objectives can be easily modeled as simpler functions of the video coding parameters by using the parametric approach of the Utility Theory [5].

A. Crispness Utility

Crispness, whose subjective nature enables it to be modeled by utility theory, is basically the perceptual similarity between the intensity edges in a digitized and compressed video, and the edges in a real-life scene, as perceived by a human viewer. The most dominating parameter, affecting the crispness of a video, is the number of bits per pixel (*bpp*) for a fixed encoder performance. 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

$$\begin{aligned} U_{crisp}(coded\ \ bits\ per\ \ pixel\ ) &= U_{crisp}\ (CBR/(CFR\ CSR)\ ) \\ &= U_{crisp}\ (CBR,CSR,CFR\ ) \end{aligned} \tag{1}$$

where *CBR, CSR, CFR* stand for Coded Bit Rate, Coded Spatial Resolution, and Coded Frame Rate, respectively. It should be noted that all video coding parameters are referred as *coded* parameters, since these factors can be viewed differently, when video is rendered on the screen of a resource limited device.

It has been shown that perceived crispness of a video increases substantially, as *bpp* value is increased [6]. However, this increase reaches to saturation after a range of values for *bpp* is exceeded. This saturation is due to the inability of the HVS to discern the difference in crispness of a picture, resulting from increasing *bpp* value beyond a certain point [7]. The crispness utility is also expected to depend on the *CSR* of the video, since perception of crispness for a given picture is related also to its resolution. In the light of the above observations it can be asserted that, the utility of crispness curve should have an exponential form, as expressed by the following formula where $c_1(CSR)$ is to be determined from subjective experiments. :

$$U_{crisp}(\mathrm{CBR},\mathrm{CSR},\mathrm{CFR}) = 1 - e^{-c_1(CSR)\frac{\mathrm{CBR}}{\mathrm{CFR\ CSR}}} \tag{2}$$

It has been shown that the perception of crispness depends on the texture content of the image under evaluation [6]. This effect can be incorporated into the proposed model by using any metric extracted from the video describing its texture and using a modified c_1 expression which is a function of this metric. Subjective tests related to the perception of crispness have been performed for different videos, having significantly varying levels of texture and the results are presented [10]. A subset of all these subjective test results can be found in Section II.D.

B. Motion Smoothness Utility

Motion-smoothness is also another subjective phenomenon, indicating the perceptual similarity of temporal motion of an event in real world, and the motion observed through the succession of video frames. The motion smoothness of a video clip can be modeled as a function of *CFR* only, if the resource constraints of the user terminals are not taken into consideration. However, the observed frame rate during playback in a user terminal will generally not be equal to the *CFR*, due to resource limitations.

Intuitively, the frame rate, at which the "observed frame rate" deviates from the original coded frame rate, should depend on the *CBR* and *CPU*. Hence, 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 video, and obviously, *CPU* of the end terminal. Thus, the functional representation for the second component of the utility function is determined as

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

Intuitively, the motion smoothness utility is expected to increase up to a point in an exponential form with increasing *CFR* and then reach to saturation (similar to the increase in crispness utility with increasing *bpp*). This effect has been demonstrated through subjective tests for different content types [8]. 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 motion smoothness utility can be modeled as follows: the exact location of the "turning point"; i.e. the frame rate at which the motion smoothness utility starts decreasing for a given bit-rate, should be determined as a function, *FR*(*CBR*). However, the rate of such a decrease in utility should differ for devices with different *CPU* capabilities.

Without losing generality, in order to simplify the formulation, only two different *CPU* configurations are utilized, while modeling the dependence of motion smoothness utility on the *CPU* of the terminal device. In other words, terminal devices, having a *CPU* clock frequency higher than a predetermined threshold value, are considered as operating in *CPU High* mode, while the ones having a lower clock speed are considered to be operating in *CPU Low* mode. The corresponding utility function for *CPU Low* is expected decrease more rapidly, compared to that of *CPU High* in the *CFR*>*FR*(*CBR*) region. Based on the reasoning above, the utility function model in (3) is proposed. Note that, "time constants" of the exponential terms, sm_{0L}, sm_{1L} sm_{0H} and sm_{1H}, for *CPU High* and *CPU Low* cases are different functions of *CBR*, yielding different increase and decrease rates at each *CBR*. It has been shown that the perceived motion jerkiness of a video depends also on the viewed content [8]. The results of the subjective tests in Section II.D indicate that for a low-motion content video, motion smoothness does not decrease, i.e. enter saturation, even at the highest bit-rate.

$$\mathrm{Usmooth(CFR,CBR,CPU)} = \left\{\begin{array}{ll} 1-e^{-sm_{0L}\,\mathrm{CFR}} & ,\mathrm{CFR} <= \mathrm{FR_L(CBR)} \\ sm_L e^{-sm_{1L}(\mathrm{CFR}-\mathrm{FR_L(CBR)})} & ,\mathrm{CFR} > \mathrm{FR_L(CBR)} \end{array}\right\} \mathrm{CPU\ Low}$$

$$\mathrm{Usmooth(CFR,CBR,CPU)} = \left\{\begin{array}{ll} 1-e^{-sm_{0H}\mathrm{CFR}} & ,\mathrm{CFR} <= \mathrm{FR_H(CBR)} \\ sm_H + 1 - e^{sm_{1H}(\mathrm{CFR}-\mathrm{FR_H(CBR)})} & ,\mathrm{CFR} > \mathrm{FR_H(CBR)} \end{array}\right\} \mathrm{CPU\ High} \tag{3}$$

$$\mathrm{FR}\ \alpha\ \frac{1}{\mathrm{CBR}} \qquad sm_{L,H} = 1 - e^{-sm_{0L,H}\mathrm{FR}_{L,H}}$$

This effect can be easily incorporated into the model by considering a video motion activity measure (e.g. MPEG-7 motion activity descriptor) and using modified *FR* and *sm* expressions that are also functions of this measure, as well as the *CBR*.

C. Content Visibility Utility

Content visibility utility is simply related to the comprehensibility and visibility of the video content with respect to its resolution and the screen size of the terminal.

The utility of the content visibility of a video clip should depend on two factors: Initial *CSR* of the video and the *screen size* of the user terminal. A video, can only be viewed partially, i.e. either cropped or down sampled, on a terminal whose screen size is smaller than the *CSR* of this video. The results of the subjective tests in the proceeding section show that cropping results in reduced user satisfaction, as expected. On the other hand, down sampling does not further reduce the satisfaction and should yield a saturated satisfaction after *CSR* exceeds the screen size. For both of these cases, the final component of the utility function is prototyped as follows:

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

Considering only the cropped case, the utility of the content visibility of a video clip is expected to increase in a similar fashion to (2) and (3), up to the point at which the spatial resolution becomes equal to the screen size of the terminal. After that point, in case of cropping, the utility is expected to decline conforming to the following equation:

$$U_{cv}(CSR, ScreenSize) = \left\{\begin{array}{ll} 1-e^{-s_1 CSR} & CSR \le ScreenSize \\ s\,e^{-s_2(CSR-ScreenSize)} & CSR > ScreenSize \end{array}\right\} \tag{4}$$

$$s = 1 - e^{-s_1 ScreenSize} \qquad s_1\ \alpha\ \frac{1}{ScreenSize} \qquad s_2\ \alpha\ \frac{1}{ScreenSize}$$

The parameters s_1 and s_2 should be both inversely proportional with the screen size of the terminal, since the increase or decrease in utility is expected to happen more abruptly in devices with smaller screen sizes. Similar to prior discussions, it should be noted that different types of content type might affect the proposed models in various ways. For instance, on close-up shots where the content fills the whole screen, the utility might decrease more suddenly, when the video frame is cropped, whereas on shots for which the scene mainly consists of a repeating pattern, such an effect may not be the case. These effects can be incorporated into the model by using a metric that defines the distribution of the content within the scene and using modified s_1, s_2 expressions which are functions of this metric.

D. Subjective Tests for Determining Utility Functions

In the next step, the utility associated with each sub-objective is determined for various video coding parameters and terminal characteristics by a series of subjective evaluation experiments. These experiments are performed in accordance with the principles stated in *ITU-R 500-11 Subjective Television Picture Assessment Standard* [8]. The tests were performed on a *Siemens Pocket LOOX 600* Personal Digital Assistant (PDA). The selected test method is the *Double Stimulus Impairment Scale (DSIS)* [8].

Figure 2(a) illustrates the results of subjective tests related to the utility of crispness for a high texture image from a sitcom.

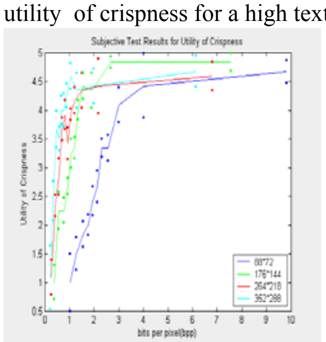

Figure 2(a): Subjective Test Results for Utility of Crispness for high textured content.

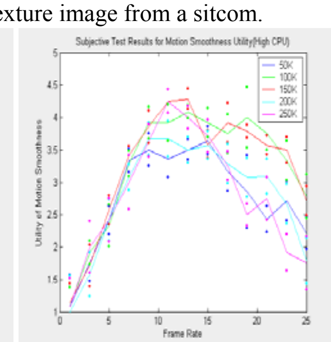

Figure 2(b): Subjective Test Results for Utility of Motion Smoothness for high motion content (High CPU)

Figures 2(b), 3(a), and 3(b) illustrate the results of subjective tests related to the utility of motion smoothness. While Figures 2(b) and 3(a) demonstrate the results for a high-motion soccer video for the high CPU and the low CPU cases respectively, Figure 3(b) shows the results for a very low motion content video, consisting of an anchorman with only limited head motion for the high CPU case. Figure 3(b) shows that even for a video encoded with 250 Kbits/s, the end terminal is able to decode this low-motion video in real time. This result is expected, since motion compensation, which is an computationally expensive phase of the decoding process (quite demanding especially for a PDA), is less utilized for such a static video, in comparison to the active video of Figures 2(b) and 3(a). Figure 4(a) illustrates the results of subjective tests related to the utility of content visibility for a sequence containing a close-up recording of a dialogue scene. Figure 4(b) shows the system recommended videos, being displayed on a typical PDA for which the simulations are performed. These images are captured from video sequences, available at `www.eee.metu.edu.tr/~alatan/adapt`.

In order to obtain the overall utility equation, it is necessary to determine the parametric functions, utilized in the proposed models for the individual utilities. Using the results of the subjective evaluation tests, these functions are obtained in terms of *CBR*, *CFR* and *CSR* by simple least squares fitting.

E. Finding Optimal Set of Encoding Parameters

After determining all the utility components, the next goal is to determine the set of encoding parameters which maximize these components for a given device. For such multiple criteria optimization problems, finding the Pareto optimal solution set is often the first step towards obtaining the optimal solution, since a *dominated solution* can not be optimal [10].

In order to determine the Pareto optimal set, a 3-D parameter space, formed from bit-rate (BR), frame rate (FR) and the spatial resolution (SR), can be sampled, so that a finite set of points (BR,FR,SR) is obtained. The values for the individual sub objectives are calculated for each point in this space for a given user terminal. Note that at this stage, each (BR,FR,SR) triplet together with the user terminal parameters is mapped into another vector $\mathbf{U}(U_C,U_{MS},U_{CV})$, composed of the utility values for the individual sub utilities. In order to determine the optimal solution, the non-dominated U vectors are selected form the Pareto optimal set of utility components [11]. This set, being Pareto optimal, contains only the vectors for which it is not possible to find another solution vector having *all* the component utilities larger than the corresponding component utilities of the member vector. The Pareto optimal set can be further refined, by discarding the solutions for which the value of any one of the component utilities is so low that the dissatisfaction associated with it impairs the judgment of the overall utility. In other words, any of the individual utilities of a member vector can not be less than a predetermined threshold (which is heuristically chosen as 20% of the maximum possible utility). Such a restriction reinforces the assumption of independence between the component utilities, since it does not allow severely impaired videos to enter the solution set.

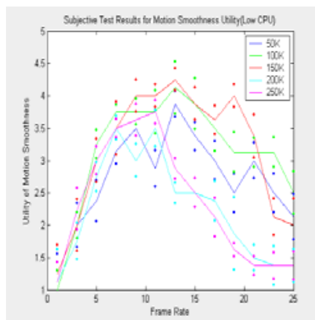

Figure 3(a): Subjective Test Results for Utility of Motion Smoothness for high motion content (Low CPU)

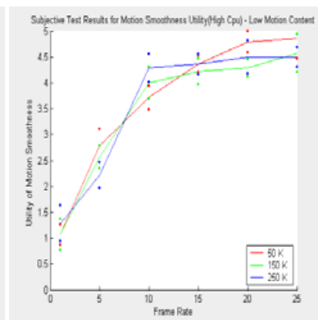

Figure 3(b): Subjective Test Results for Utility of Motion Smoothness for low motion content (High CPU)

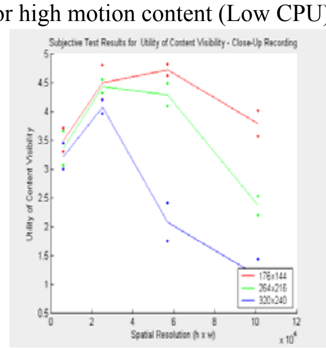

Figure 4(a): Subjective Test Results for Utility of Content Visibility Close-Up Recording

Figure 4(b): Compaq IPAQ displaying video at recommended parameters.

Once the Pareto optimal set is determined, the effect of finite channel capacity can be considered. For a given maximum channel capacity, the members of the Pareto optimal set having associated bit-rate (BR) values higher than this capacity are discarded from the Pareto optimal set. The remaining members are all suited for transmission through the given channel. In order to choose a specific solution from the remaining members, an additional figure of merit needs to be selected. In the simulations presented in the following section the solution having the highest associated bit-rate, i.e. the bit rate closest to the channel capacity, has been chosen so as to utilize the channel to the fullest extent. Choosing another criterion to select a specific member of the set, such as having the highest motion smoothness utility or highest crispness utility are equally valid.

III. SIMULATIONS

A. Effects of Finite Channel Capacity

3D parameter space is sampled into 6000 discrete points. Then, the Pareto optimal solution set is obtained by using the procedure outlined in Section II.E. If there are no restrictions on the channel capacity, the Pareto optimal set contains 1234 members. Figure 5 shows the members of the Pareto optimal set when the maximum channel capacity is restricted to 75 Kbits/s. 92 members have associated bit rates lower than the specified capacity. The marked solution is chosen, as it has the highest associated bit-rate, i.e. the bit rate closest to the channel capacity as specified in the previous section.

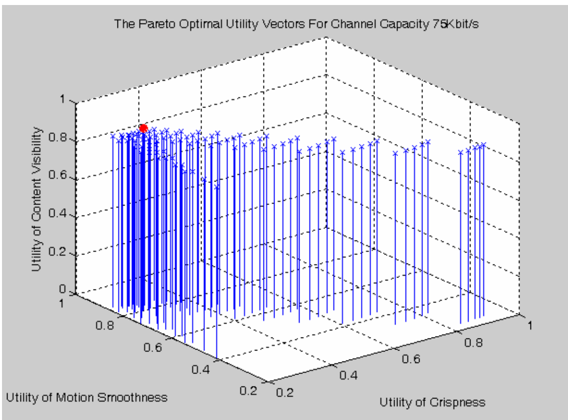

Figure 5 : The Pareto optimal utility vectors for a channel with capacity 75 Kbits/s

When the maximum channel capacity is increased to 100 Kbits/s, 211 members out of the initial 1234 remain in the solution set. The solution that has the highest associated bit-rate is once again marked in the plot. The total execution time required to obtain the Pareto optimal set is slightly less than 2 seconds in a Intel Pentium III laptop computer.

IV. CONCLUSIONS

The main contribution of this paper is inclusion of transmission channel capacity into the previously proposed subjective utility models for user viewing satisfaction in resource limited devices. It has been shown that using the maximum channel capacity as a restriction metric, certain members of the Pareto optimal solution set can eliminated such that the remaining members are suitable for transmission through the given channel. Once the reduced Pareto optimal set is obtained, an additional figure of merit can be used to pick a single solution from this set depending the application scenario. Finally, it should be stated the proposed utility models for video adaptation scenario is quite generic in the sense that different content types or channel capacity restrictions, can still be incorporated into these models.

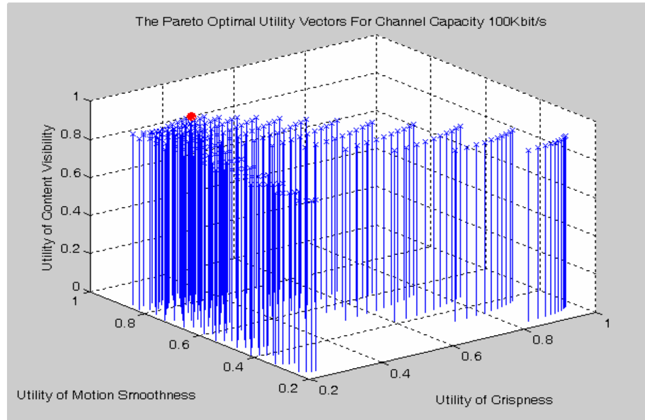

Figure 6 : The Pareto optimal utility vectors for a channel with capacity 100 Kbits/s

V. REFERENCES

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

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

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

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

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

[6] M. Ardito, M. Barbero, M. Stroppiana, M. Visca, "Compression and Quality," Proceedings of the International Workshop on HDTV, 1994

[7] K. Boff, L. Kaufman, & J. Thomas (Ed.), "Handbook of Perception and Human Performance : Sensory Processes and Perception," John Wiley and Sons Inc, 1986.

[8] Ronnie T. Apteker, James A. Fisher, Valentin S. Kisimov, and Hanoch Neishlos, "Video acceptability and frame rate," IEEE Multimedia Volume: 2 , Issue: 3 , Fall 1995 Pages:32 – 40

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

[10] Ö. D. Önür, A. A. Alatan, "Optimal Video Adaptation for Resource Constrained Mobile Devices Based on Utility Theory," WIAMIS 2004, Portugal.

[11] Ralph E. Steuer, "Multiple Criteria Optimization: Theory, Computation and Application," John Wiley and Sons Inc, 1985