A FRAMEWORK FOR PERSONALIZED UTILITY-AWARE IP-BASED MULTIMEDIA CONSUMPTION

Martin Prangl and Hermann Hellwagner

Dept. of Information Technology
Klagenfurt University, Austria

ABSTRACT

Providing transparent and augmented use of multimedia content across a wide range of networks and devices is still a challenging task within the multimedia research community. Multimedia adaptation was figured out as a core concept to overcome this issue. Most multimedia adaptation engines for providing Universal Multimedia Access (UMA) scale the content under consideration of terminal capabilities and resource constraints but do not really consider individual user preferences. This paper introduces an adaptive multimedia framework which offers the user a personalized content variation for satisfying his/her individual utility preferences.

1. INTRODUCTION

The delivery of multimedia content over best effort networks (like the Internet) through various types of communication channels (WLAN, UMTS, Cable etc.), becomes more and more important. Modern terminals like PDAs or mobile phones enable the community to receive multimedia content every time and everywhere. In order to achieve this UMA vision content adaptation is necessary to meet the terminal capabilities, network characteristics and user requirements. But how should the content be adapted to provide the maximum utility to the end user? Most adaptive multimedia frameworks are adapting the content by simple frame dropping, requantization or rescaling of a video, for instance, according to given resource limitations and terminal capabilities.

Such systems address the terminal as the end point of the adaptation chain and not the consuming user itself. However, an adapted audiovisual content is a combination of the adapted elementary stream variations. Both, video and audio can be adapted in several ways, e.g., the video can be modified in the temporal, in the spatial, or in the SNR domain. The audio part can be modified by reducing the number of audio channels, reducing the sample frequency or reducing the audio bitrate. Therefore, in our opinion, the question "How to adapt multimedia data in order to provide the best user perceived utility?" is of central relevance and needs to be addressed. The basis of our answer forms the development of a generic cross-modal utility model presented in [1] and a hybrid recommender system, which tries to configure the utility model optimally for the individual user, presented in [2]. We integrated these approaches into a powerful multimedia framework which we call Personal Live Video Server (PLVS).

The remainder of this paper is organized as follows. Section 2 gives an overview of the main features of PLVS. Section 3 explains the architecture of our system. In section 4 we summarize the results of a detailed evaluation of the framework's success regarding the individual user satisfaction.

2. SYSTEM FEATURES

Our proposed framework aims to offer a personalized content variation to the individual user and his/her environment. Therefore it considers terminal specific properties like the type of the terminal (notebook, PDA, mobile, ...) and its properties (network connection, display size, etc.). Furthermore, it captures user specific properties like favorite content types (genres) and age as well as natural environment conditions such as brightness and loudness. Together with the information about the original content like genre and low level information (resolution, frame rate, audio sample rate, number of audio channels, etc.) as well as the feedback of the utility impression of other users, an adaptation decision for the current request is calculated. All user, terminal, and content specific information can be registered by a convenient Web interface. Additionally, user and terminal specific information can be uploaded using standardized MPEG-21 Usage Environment Descriptors (UEDs) [3].

PLVS enables the user to consume video on demand as well as live television and radio streams. Live television and live radio channel consumption is enabled by a DVB[1] receiver card attached to the system. The personal user information is reused for a realization of a multimedia content recommender which suggests the active user some videos of his/her interest.

3. SYSTEM ARCHITECTURE

The system architecture of PLVS can mainly be divided into four modules as shown in Fig. 1 and briefly introduced as

[1] http://dvb.org

1-4244-0992-6/07/$25.00 ©2007 IEEE

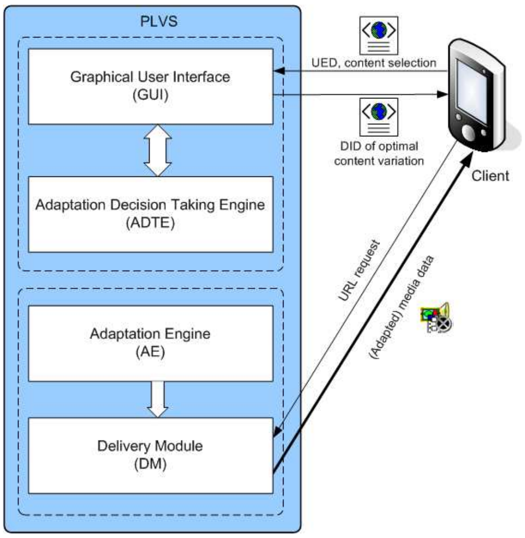

Fig. 1. System architecture of PLVS

follows.

3.1. Graphical User Interface (GUI)

The GUI is implemented as an online portal. It offers the user a register form, where he/she is able to submit his/her preferences. Furthermore, the user is able to register his/her terminal properties and environment conditions and upload some content in his/her personal repository. The upload of a user and terminal specific MPEG-21 UED is provided as well. All user data and content specific metadata are stored in a database. The user's consumption history including his/her utility impressions are maintained in the database as well. The content can be consumed directly via the GUI. This functionality is enabled by including the VLC ActiveX plug-in[2] or the Macromedia flash player[3]. Optionally, the user is able to consume the content by using an external media player. The electronic program guide (EPG) which provides information like title, duration, and abstract of the available TV channels, is presented by the GUI as well.

3.2. Adaptation Decision Taking Engine (ADTE)

If the user requests a specific content, the adaptation decision taking engine (ADTE) is invoked. Its task is to select the optimum adaptation parameters for the content adaptation such that the given bandwidth limit is not exceeded, the terminal capabilities (available codecs, spatial resolution, ...) are fulfilled, and the user experience is maximized. In order to achieve this target, a hybrid recommender system [2] is invoked to configure our cross-modal utility model [1]. Based on the individually configured utility model, the ADTE is calculating the optimum adaptation parameters based on optimization algorithms presented in [4]. The calculation requires additional information about the properties of the original content stream including the genre information. This information is extracted/defined by the upload process into the user's content repository. In case of a live TV stream, the genre information is taken form the online program guide called *hoerzu*[4]. Finally, the ADTE generates an individual URL on which the client's player (or plug-in) can access the adapted content. This URL is containing the information on how to adapt the content and is optionally encapsulated together with metadata of the adapted content variation (frame rate, spatial resolution, etc. as well as high level metadata like title, genre, and abstract) into an MPEG-21 DID [5], which is accessible for the client w.r.t. interoperable access. A simplified example of the DID including the URL is given in Fig. 2.

```
<did:DIDL>
 <did:Item>
  <did:Component>
   <did:Descriptor>
    <did:Statement mimeType="text/xml">
     <Mpeg7>
       <!-- metadata information here -->
     </Mpeg7>
    </did:Statement>
   </did:Descriptor>
   <did:Resource mimeType="video/mpeg"
     href="http://plvs-itec.uni-klu.ac.at/
        examplevideo.mpg?videocodec=mpeg4&audiocodec=mp3&
        resolution=320x200&framerate=15&
        audiochannels=2$samplerate=44100"/>
  </did:Component>
 </did:Item>
</did:DIDL>
```

Fig. 2. MPEG-21 DID of a content variation including metadata and URL, created for an individual request (simplified)

3.3. Adaptation Engine (AE)

The adaptation engine (AE) is responsible for adapting the original audiovisual content into a degraded variation based on the decision of the ADTE. The adapted content variation is encoded by a codec which is supported by the player on the client device (decided by the ADTE). The implementation of the AE is based on a powerful multimedia library called *ffmpeg*[5] which supports various encoders including the *3gp* video codec and the *amr* audio codec for mobile phones. The AE is adapting the requested media stream on the fly. Its output is forwarded to the delivery module.

3.4. Delivery Module (DM)

The delivery module (DM) is delivering the adapted media stream to the client. For this task, we make use of the TCP-

[2] http://www.videolan.org
[3] http://www.macromedia.com
[4] http://www.hoerzu.de
[5] http://ffmpeg.mplayerhq.hu

based HTTP protocol. The reason is that we want to avoid problems caused by firewalls, NAT routers or proxy servers which are nowadays widespread at clients' sites. Using the UDP-based RTP protocol would cause problems in this context. If the media player at the client side opens the prepared HTTP link, the DM parses the adaptation parameters from the received URL and configures the AE accurately. Finally, it starts pulling the adapted content from the AE and forwards it to the media player for rendering.

After the user has consumed the content, he/she is asked to rate it based on his/her individual utility impression. This feedback influences the adaptation decision for requests of similar users under similar terminal and environment conditions. Note that the lossless TCP connection avoids content degradation by the delivery process and consequently ensures reliable feedback data. For more information about the proposed recommendation technique and the user friendly feedback mechanism the reader is referred to [2].

4. EXPERIMENTAL RESULTS

In order to evaluate the success of the proposed utility-aware multimedia system, we performed subjective tests based on four audio-visual scenes of different genres which have been recorded from a digital television stream (DVB). An action genre was covered by a scene of *Stargate*, a soccer game *(Rapid Vienna vs. Juventus Turin)* has been taken for sports, the third content was a talking head news clip taken from *n-tv*, and a scene of *Universum* showing an octopus was our choice for a representation of a documentation clip. Based on this original content we produced 84 audio-visual output variations in total by applying three bandwidth limitation constraints to each content type. The duration of all content variations was set to 10 seconds. The Subjective Mean Opinion Score (MOS) values for all variations were obtained by 30 test persons, all students from Klagenfurt University and non-experts. For the MOS feedback strategy we decided to use the Absolute Category Rating (ACR) according to ITU.T Rec. P.910 which defines an eleven grade scale (from bad to excellent). It took 25 minutes on average to execute one subjective test program and the critical limit of 30 minutes was never exceeded. We assured the same external environment regarding brightness and noisiness for each test person by executing the tests in the Usability Lab of Klagenfurt University. The test clips were shown on a TFT monitor placed 50 cm in front of the user, headphones were used to provide reliable audio conditions. The MOS rating values were added step-by-step to the database in the same order as in the actual subjective tests. Each time a new user was added to the database we used the recommender to get rating predictions for this user in advance. This allowed us to monitor a mean absolute difference (delta MOS) by comparing recommender predictions and real rating values given by the test candidates. As a concluding result, we found that the average MOS error is decreasing with increasing number of users and recommendations for all four content types. This indicates that the reliability of the predicted optimal adaptation decision increases with the number of known ratings. For a more detailed description of the evaluation the reader is referred to [6].

5. CONCLUSION

We presented an overview of our utility aware adaptive multimedia framework, which aims to offer UMA for the individual user. The system creates a personalized adapted content variation for his/her utility satisfaction. We introduced the main features of the proposed system, including the possibility of live consumption of TV content. We gave an overview of the system architecture, where we briefly explained the functionality of each module. Finally, we reported results of a detailed evaluation based on subjective tests. We found that the system's adaptation decision taking engine offers reliable results: its success providing adapted content to satisfy the individual user's preferences and constraints is increasing with the number of feedback ratings.

6. REFERENCES

[1] M. Prangl, H. Hellwagner, and T. Szkaliczki, "A Semantic-based Multi-modal Utility Approach For Multimedia Adaptation", *Proc. 7th Int'l. Workshop on Image Analysis for Multimedia Services (WIAMIS)*, April 2006.

[2] M. Prangl, R. Bachlechner, and H. Hellwagner, "A Hybrid Recommender Strategy for Personalized Utility-based Cross-modal Multimedia Adaptation", *Proc. IEEE Int'l. Conf. on Multimedia and Expo (ICME)*, July 2007.

[3] A. Vetro and C. Timmerer, "Digital Item Adaptation: Overview of Standardization and Research Activities", *IEEE Trans. on Multimedia*, vol. 7, no. 3, June 2005.

[4] M. Prangl, H. Hellwagner, and T. Szkaliczki, "Fast Adaptation Decision Taking for Cross-modal Multimedia Content Adaptation", *Proc. IEEE Int'l. Conf. on Multimedia and Expo (ICME)*, July 2006.

[5] M. Prangl, C. Timmerer, and H. Hellwagner, "Real-time DVB-based MPEG-21 Digital Item Adaptation for Live Universal Multimedia Access", *Proc. 48th Int'l. Symposium ELMAR-2006*, Zadar, Croatia, 2006.

[6] M. Prangl, T. Szkaliczki, and H. Hellwagner, "A Framework for Utility-based Multimedia Adaptation", *IEEE Trans. on Circuits and Systems for Video Technology*, 2007 (accepted for publication).