Abstract—Virtual collaboration concept allows remotely located partners to meet in a virtual environment using the state of the art communication and digital multimedia technologies. A virtual collaboration system is being developed under VISNET II Network of Excellence of the IST FP6 programme. This paper elaborates on a virtual collaboration scenario and proposes content adaptation technologies to fulfill needs of all participants. The proposed content adaptation techniques are based on scalable video coding. The content adaptation based on user preferences, profiles for usage, terminal and network capabilities are discussed. Initial experimental results have been proved to be effective and productive for a virtual office environment.

Index Terms—Virtual Collaboration, Content Adaptation, Scalable Video Coding, H.264.

I. INTRODUCTION

The success of the Internet and mobile systems has motivated the development of various enhanced-capacity fixed and wireless networking technologies (e.g., 3G, WLAN, WiMAX, broadband Internet etc). The services supported by such networks help foster the vision of being connected at anywhere, anytime and with any device for pervasive media applications. Virtual collaboration system (VCS) allows remotely located partners to meet in a virtual environment using the state of the art communication and audiovisual technologies. In a virtual office environment, all remotely located partners will feel the sensation of being in a single room regardless of their true geographical location. Fig. 1 demonstrates a context diagram of a virtual collaboration scenario.

This paper presents content adaptation mechanisms for the VCS described by VISNET II Network of Excellence (NoE) [1] of the IST FP6 programme. The VCS is currently being designed by the NoE partners in order to achieve cross-theme integration by bringing together the work performed in the three VISNET II themes, namely video coding, audiovisual media processing and security, using Virtual Collaboration as an integration driver. In the following sections, the discussion will focus on the content adaptation as applicable to virtual collaboration scenarios.

In the virtual collaboration scenario as shown in Fig. 1, large fixed-terminal acts as the main control/command point and serves for group of co-located users. This may be the headquarters of the organization and consists of communication terminals, shared desk spaces, displays and various user interaction devices to collaborate with remotely located partners. The remotely located users with a small, fixed terminal will act as the local contact and provide the local information. Mobile units (distribution, surveying, marketing, patrolling, etc) of the organization will use mobile terminals such as mobile phones and PDAs to collaborate with the headquarters. The top level system decomposition of the VCS with a terminal for each group of co-located users is given in Fig. 2. The gateway associated with each terminal may be a part of the terminal or the interconnection depending on the implementation. For example, it is likely to be included in the large multi-user terminal but separated by a wireless link from the small mobile terminal. Remote services provide persistent storage and workflows, authorization and awareness of other users’ connectivity.

All remotely located collaborators use range of terminal devices with different capabilities. For example small mobile terminals may have different display sizes, stream handling capabilities (speech, video, scalable video, etc), processing
power and memory capacity compared to large fixed terminals in the headquarters, which provides rich multimedia in an unrestricted environment (large displays, broadband communication links with high processing power).

The mismatches between the content properties and several network and/or device-centric features, as well as diverse user preferences in a virtual collaboration scenario call for efficient multimedia delivery systems, featuring effective content adaptation mechanisms [2]. The content adaptation mechanisms will improve the subjective quality of the user experience in a virtual office scenario. Furthermore, this improves the effectiveness and overall productivity of the collaborative activities. In general, content adaptation concept has been addressed in literature with the theme of the universal multimedia access (UMA) [3][4]. Several strategies have been developed for UMA, which are based on the multimedia content adaptation techniques using the context specifications and descriptions defined in Part-7: Digital Item Adaptation (DIA) of the MPEG-21 standard [5][6][7].

An effective way of performing content adaptation is to utilise transcoding operations in the networks. Transcoding is particularly needed when compressed media streams traverse heterogeneous networks. In such cases, a number of content-specific properties of the coded multimedia information require adaptation to new conditions imposed by the different networks and/or terminals to retain an acceptable level of service quality. Network based adaptation mechanisms can be employed at the edges or other strategic locations of different networks (at terminal gateways, see Fig. 2), using a fixed-location content adaptation gateway, node or proxy as in conventional networking strategies ([8][9]). Alternatively, content adaptation through transcoding can be performed dynamically where and whenever needed using active networking technologies [10]. In a virtual collaboration scenario, content can be adapted using transcoding based on two main approaches depending on the scalability of the multimedia content. If the original content is scalable, a sub-stream can be extracted depending on the terminal/network capabilities and user preference. In the second approach, the non-scalable content will be reprocessed to suit user preferences and capabilities. However, transcoding of non-scalable content requires high computational power at network nodes and incurs heavy delays. Therefore, this approach is not considered for the VCS. Hence, the rest of the paper focuses on transcoding of scalable content in a virtual collaboration scenario.

This paper proposes content adaptation mechanisms to be used in a virtual collaboration scenario based on scalable video coding (SVC) approach. SVC is a feasible video adaptation technique that fits within the MPEG-21 DIA architecture. The scalable video coding in a VCS allows encoding the content in several scalable layers. Number of layers is dependent on the scalability space required by the application scenario. SVC approach provides backward compatibility for conventional multimedia such as speech and video, while providing the adaptation for different networks and/or terminal capabilities and different user preferences. Hence, the heterogeneity of multimedia communication using different terminals in a virtual office environment can be addressed using this proposed SVC based content adaptation mechanism. The emerging of real-time scalable video codecs with high compression efficiency, availability of high processing power and ever increasing network bandwidth will facilitate real-time content generation and adaptation based on this proposed SVC approach.

The proposed SVC based adaptation for the VCS is described in Section II. Section III elaborates on adaptation mechanisms based on the conditions imposed by user preferences, profiles for usage, terminal and network capabilities. Furthermore the same section provides examples of application specific scenarios such as bitrate and user preference based adaptations. Section IV concludes the paper.

Fig. 2 describes the virtual collaboration scenario based on the conditions imposed by user preferences, profiles for usage, terminal and network capabilities. Furthermore the same section provides examples of application specific scenarios such as bitrate and user preference based adaptations. Section IV concludes the paper.

II. CONTENT ADAPTATION USING SVC APPROACH

Fig. 3 describes the virtual collaboration scenario incorporating the proposed SVC based content adaptation scheme. The content adaptation based on SVC approach consists of two main modules namely SVC codec for scalable content generation and the adaptation module as shown in Fig.
3. The adaptation module consists of an adaptation decision engine (ADE) and adaptation engine (AE). The ADE will take all decisions based on the network capacity, terminal capabilities and user preferences and use the scalable content available at the gateway. The AE implements the decision made by ADE. The SVC codec resides in major user terminals (large user terminals with co-located users and other fixed user terminals with single users). Even mobile user terminals will join to provide scalable multimedia content, depending on their terminal capabilities. The adaptation module is residing in the terminal gateway of the particular user or the terminal itself.

The interaction of the SVC codec and the terminal gateway is demonstrated in Fig. 4. The required scalability space is utilised during a virtual collaboration session, including audio/video and documents. The objective of this section is to discuss video adaptation mechanisms proposed for virtual collaboration application to cater for the heterogeneity of the channel conditions, user preferences and terminal capabilities.

A. Adaptation Based on User Preferences and Profiles

It is more likely that each user of virtual collaboration system would have own audio and video preferences. For instance, a user may wish to select a specific area which draws his/her main attention in visual content. Thus, he or she may want to access a part of the video scene based on his/her selection. In addition to this or in a totally isolated situation, the terminal that the user is using may have a restricted display capability with lower resolution than that of the originally encoded content. Moreover, the access network that the user is connected to may not be able to support elaborate visual information transfer due to bandwidth limitations and/or other channel specific characteristics.

All of these add up to the profiling of a use case for this particular user, and the different display capabilities, attention area selection preferences, access network based features etc provide the necessary context elements for this use case. Therefore, more user centric adaptation options must be available for serving each individual’s requirements.

Fig. 3: Integration of content adaptation methodology in VCS

Fig. 4: Interaction between the SVC and the adaptation module

The interaction of the SVC codec and the terminal gateway is demonstrated in Fig. 4. The required scalability space is dependent on the application scenario, terminal capabilities of remotely located users, user preferences and the type of access network, where remote users are connected. After in-depth analysis of the heterogeneity of the prevailing virtual collaboration scenario, appropriate scalable content can be generated using the SVC codec available at fixed terminal with single/multiple users. All or part of the generated scalable multimedia content will be transferred to the terminal gateway depending on the capacity of the network (see Fig. 3). The rest of the paper discusses the functionalities of the adaptation module with respect to network/terminal capabilities and user preferences.

III. CONTENT ADAPTATION FOR VIRTUAL OFFICE ENVIRONMENT

Virtual collaboration is an application which relies heavily on digital media. Therefore, the effectiveness of this application can be enhanced by adapting the relevant media items for the usage environment. Adaptation possibilities spread across whole spectrum of digital media items that are utilised during a virtual collaboration session, including audio/video and documents. The objective of this section is to discuss video adaptation mechanisms proposed for virtual collaboration application to cater for the heterogeneity of the channel conditions, user preferences and terminal capabilities.

Fig. 5. User centric adaptation mechanism for virtual collaboration application

In the virtual office application, one of the more demanding user centric adaptation scenarios is delivering a cropped view of a particular attention area selected by the user. This area is more subjective and situation driven than an automatic attention area detection technique may be able to detect. In order to address these conditions, the adaptation technique illustrated in Fig. 5 is proposed. Unique to the proposed adaptation mechanism is the provision of a user interaction service, through which the user can specify a selected ROI. This information is sent in an ROI Selection message to the ADE which is implemented as a part of the relevant gateway as shown in the Fig. 5 or as a stand-alone network element. Based on the ROI definition and other context information
defining the capabilities of network and terminal etc., the ADE determines the nature of adaptation needed after processing the ROI descriptor. This adaptation decision is used for configuring the Adaptation Engine (AE) which performs actual adaptation.

The AE utilises the provisions provided by H.264 SVC [11] extension for ROI based adaptation using Flexible Macroblock Ordering (FMO). This technique is formally identified as interactive ROI (IROI) scalability [12][13]. An AE can utilise the IROI scalability to extract a sub-stream that defines the ROI [14]. Fig. 6 illustrates the perceptual quality comparison of downsampling and IROI based adaptation of CIF sequences to match QCIF resolution. It is clearly evident that cropping based adaptation is more attractive when the preservation of fine details of the attention area is more important.

In addition to the user driven ROI feedback signalling, the ROI information is used when there is a necessity of adapting the content to meet usage environment constraints. For example, if there is a need to reduce the aspect ratio, then the adaptation operation can be performed taking the ROI information into consideration; or if it is necessary to reduce the bitrate, the ADE may decide that the best way of achieving that is by cropping the image according to the ROI information. All of these are determined after processing a number of context descriptors, which in return describe the user defined ROI and other constraints, such as QoS, terminal capabilities, access network capabilities, usage environment etc.

B. Adaptation Based on Terminal and Network Capabilities

The use of various devices (e.g., PCs, PDAs, laptops, other large/small scale fixed/mobile collaboration terminals etc) in a virtual collaboration system often demands the adaptation of content format during the exchange of information between collaborating partners due to their diverse terminal capabilities. These capabilities may sometimes vary from one terminal to another vastly, in terms of the type of the codec used, amount of memory storage, supported processing power, available communication and/or user-interface modules (e.g., video, speech, audio, text facilities etc), the size of display etc. In most cases, adaptation needs to be performed during the access of rich content information. It is evident that the user terminal with the least number of capabilities will be the limiting factor on others’ communication and collaboration activities. To avoid the fact that such limitation affects the overall QoS assured and the quality of user experience (QoE) expected from this scenario, a content adaptation engine can be deployed to act as a guarantee of interoperability between various user terminals with diverse capabilities. In such a scenario for instance, a mobile terminal can be regarded as a power-limited device due to battery life and processing power restrictions, and hence is conventionally built with a small (low-resolution) screen. The high resolution of video can thus be re-sized to fit adequately on such a terminal with small display size without any information loss, which in return will enhance both the QoS and QoE. If reducing the original size of video is not providing the best results, then selecting a particular attention area of user’s choice (i.e., ROI selection) can be a suitable way of adapting the content for improved interoperability.

There can also be potential occasions where the user terminal does not support, say the video element or it cannot match the format of the originally compressed video and/or audio. In such situations, adaptation can take place to provide conversions between different content coding standards as well as selecting one and dropping the other content item or perform mode exchanging (i.e., transmoding) based on the supported feature. Transmoding engines can perform video to sound/text, to images and/or to graphic conversions [15].

The availability of network bandwidth for different users is not fixed in a virtual office scenario. For example, a PDA user may need to access the interaction through a wireless connection. The bandwidth availability over a wireless network is limited and time varying. Hence the user can request low bitrate content from the adaptation module located in the terminal gateway. The adaptation module can decide and deliver a down-sampled version of the collaborated content using available scalable multimedia content. The spatially or temporally down-sampled content can be extracted from the scalable stream to obtain low bitrate streams. Furthermore incremental refinement nature of the fidelity scalability tools supports termination of enhancement layer packets at supported rate points [16]. Therefore, this serves as a powerful tool to adjust the bitrate by truncating an enhancement layer packet. A different approach of achieving bitrate adoption is to use the SVC’s ROI scalability feature. The advantage of this technique is that the quality of the attention area can be preserved while easing the channel bitrate
Video if compatible displays and synthesizers are available. This section analyses a scenario of bitrate adaptation based on several spatial layers. This experiment used the test sequences, namely ‘Orbi’ and ‘Interview’ obtained using the depth-range camera. Both sequences contain a colour sequence and associated per-pixel depth information. This depth image in conjunction with the colour image can be used to render stereoscopic video material at the receiver side [17]. These sequences were selected to represent a use case, where stereoscopic video can be generated at terminals with high capacity fixed terminals including stereoscopic capture and display devices. In general the use of stereoscopic video will enhance the effectiveness of the collaboration activity between the partners. The H.264/SVC codec (JSVM 7.12.1) was used to obtain a scalable video bit-stream. The H.264/SVC supports spatial, temporal and quality scalability for video [11]. The colour sequences of different formats: QCIF, CIF and 4CIF were coded at three different layers using the H.264/SVC layered architecture. The corresponding depth image, which is in 4CIF format was coded as the forth layer [18]. The basic encoding parameters are: 125 frames, IPPP… sequence format, 30 frames/s original frame rate, a single reference frame, variable length coding (VLC), 32 pixel search range and no error resilience. The QP value of 30 was used at each layer for encoding. In order to remove the redundancies between inter views inter layer prediction was enabled among the layers. Table I shows the resultant bitrates for each coded layer.

<table>
<thead>
<tr>
<th>Sequence &amp; Format</th>
<th>Orbi</th>
<th>Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colour: QCIF</td>
<td>80</td>
<td>35.79</td>
</tr>
<tr>
<td>Colour: CIF</td>
<td>296</td>
<td>37.66</td>
</tr>
<tr>
<td>Colour: 4CIF</td>
<td>1078</td>
<td>37.99</td>
</tr>
<tr>
<td>Depth: 4CIF</td>
<td>1655</td>
<td>40.05</td>
</tr>
</tbody>
</table>

According to Table I, a QCIF format video, which has the lowest bitrate can be adapted to collaborate with a mobile user connected via a UMTS link (384 kbits/s), whereas a user with a laptop with WLAN connectivity may access the 4CIF format video. The fixed terminal users with wired connectivity may access the 4CIF format video and even experience stereoscopic video. The fixed terminal users with wired connectivity may access the 4CIF format video. The lowest bitrate can be adapted to collaborate with a mobile user on several spatial layers. This experiment used the test sequences, namely ‘Orbi’ and ‘Interview’ obtained using the depth-range camera. Both sequences contain a colour sequence and associated per-pixel depth information. This depth image in conjunction with the colour image can be used to render stereoscopic video material at the receiver side [17]. These sequences were selected to represent a use case, where stereoscopic video can be generated at terminals with high capacity fixed terminals including stereoscopic capture and display devices. In general the use of stereoscopic video will enhance the effectiveness of the collaboration activity between the partners. The H.264/SVC codec (JSVM 7.12.1) was used to obtain a scalable video bit-stream. The H.264/SVC supports spatial, temporal and quality scalability for video [11]. The colour sequences of different formats: QCIF, CIF and 4CIF were coded at three different layers using the H.264/SVC layered architecture. The corresponding depth image, which is in 4CIF format was coded as the forth layer [18]. The basic encoding parameters are: 125 frames, IPPP… sequence format, 30 frames/s original frame rate, a single reference frame, variable length coding (VLC), 32 pixel search range and no error resilience. The QP value of 30 was used at each layer for encoding. In order to remove the redundancies between inter views inter layer prediction was enabled among the layers. Table I shows the resultant bitrates for each coded layer.

### Table I

<table>
<thead>
<tr>
<th>Sequence &amp; Format</th>
<th>Bitrate (kbits/s)</th>
<th>PSNR (dB)</th>
<th>Bitrate (kbits/s)</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colour: QCIF</td>
<td>80</td>
<td>35.79</td>
<td>67</td>
<td>34.35</td>
</tr>
<tr>
<td>Colour: CIF</td>
<td>296</td>
<td>37.66</td>
<td>241</td>
<td>35.46</td>
</tr>
<tr>
<td>Colour: 4CIF</td>
<td>1078</td>
<td>37.99</td>
<td>777</td>
<td>37.13</td>
</tr>
<tr>
<td>Depth: 4CIF</td>
<td>1655</td>
<td>40.05</td>
<td>1220</td>
<td>42.30</td>
</tr>
</tbody>
</table>

**IV. CONCLUSION**

This paper elaborates on a Virtual Collaboration System (VCS) which is being developed under VISNET II and the complexity arises with the heterogeneity of the system due to use of various terminals, different access networks with diverse characteristics, user preferences and profiles. In order to address diverse perspectives of the system, we also propose content adaptation methodologies. The objective of the proposed adaptation mechanisms is to maximise the effectiveness of collaboration sessions by fulfilling the needs of all collaborators. The content adaptation based on user preferences, profiles for usage, terminal and network capabilities are described and results are presented for specific adaptation scenarios. Initial experimental results have been proved to be effective and productive for a virtual office environment. Most of the presented tools are still being developed and will be tested in the VCS.

**ACKNOWLEDGEMENTS**

The work presented was developed within VISNET II, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme. Authors would like to thank Dr Banu Gunel of I-Lab/CCSR, University of Surrey for her review comments.

**REFERENCES**