Resource-Driven Content Adaptation*

Yung-Hsiang Lu, David S. Ebert, and Edward J Delp
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana
{yunglu, ebertd, ace} @purdue.edu

ABSTRACT

Recent trends have created new challenges in the presentation of multimedia information. First, large, high-resolution video displays are increasingly popular. Meanwhile, many mobile devices, such as PDAs and mobile telephones, can display images and videos on small screens. One obvious issue is that content designed for a large display is inappropriate for a small display. Moreover, wireless bandwidth and battery lifetime are precious resources for mobile devices. In order to provide useful content across systems with different resources, we propose "resource-driven content adaptation" by augmenting the content with metadata that can be used to display or render the content based on the available resources. We are investigating several problems related to resource-driven content adaptation. These include: adaptation of the presented content based on available resources— display resolution, bandwidth, processor speed, quality of services, and energy. Content adaptation may add or remove information based on available resources. Adaptive content can utilize resources more effectively but also present challenges in resource management, content creation, transmission, and user perception.

1. INTRODUCTION

Most digital media, such as digital images and video, are displayed on desktop systems. Recently, several trends have created new challenges in presenting digital content. Many people carry mobile devices with small displays (typically smaller than 3 inches with 640×480 pixel resolution). Meanwhile, high-resolution displays have become increasingly popular. A single display can have more than 2560×1600 pixels, or 13 times more pixels than a handheld display. With this large difference in display capability, the content in many applications needs to be adapted to the display and the user environment. One simple approach is to down sample the image for a smaller display or to replicate pixels to fill a larger display. This is rarely acceptable in that the images have many spatial and/or temporal artifacts. It is possible to use sophisticated methods based on apriori statistical information for shrinking or expanding the images. Even though it may produce acceptable displayed images, this approach does not take full advantage of the large display, nor does it ensure that the crucial information is presented on the smaller display and it may not use all information available about the content or the display environment. Another trend is that devices have widely available network connections, either wired or wireless. Finally, as more functionalities are integrated into mobile devices (such as mobile phones, PDAs, and laptop computers), resource (battery life, bandwidth, computation, storage) conservation becomes a crucial issue. We believe that the ability to adapt visual content to the display, user, and environment is important to many mobile collaborative applications, including education, crisis coordination and response, and video conferencing.

At Purdue University we have a long term project where we are developing methods to address the problem of resource-driven adaptation of visual content (both still images and video). The purpose of this paper is to describe a "snapshot" of the current status of our project. We are addressing the above challenges by augmenting the multimedia content with metadata, stored in a central database, that can be used to display or render the content in a manner that is sensitive to the resources available. These resources include the display, environment, performance requirements, and resource constraints. We are investigating novel methods, based on traditional graphics rendering, non-photorealistic rendering (NPR), and image based rendering, to display an image in a way that uses the capability of the display. We feel that content

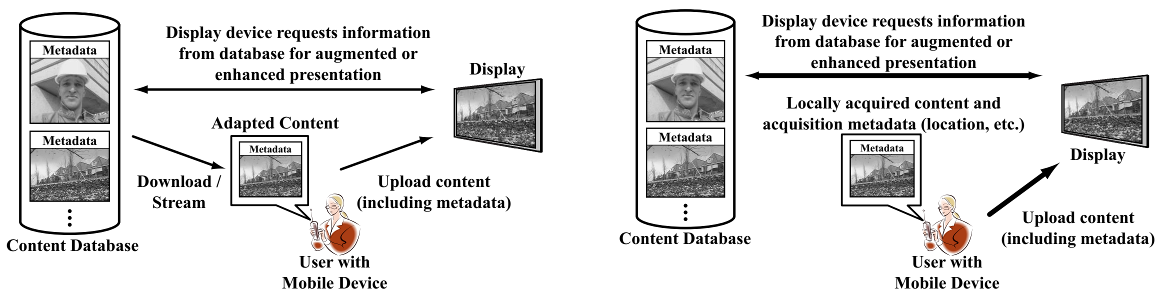
*This work was partially supported by a grant from the Indiana 21st Century Research and Technology Fund.

adaptation is one of the last barriers that needs to be addressed for "true" collaboration in heterogeneous environments. The system includes the following essential components:

- Development of novel image rendering methods for resource-driven content adaptation.
- Design of a framework so that metadata can be generated automatically for computer generated content and for content acquired through other media, such as cameras and camcorders.
- Design of control algorithms for the display that can adjust the content based on the capability of the display and the resources in the operating environment.
- Design of methods for novel exploitation of the resources in the environment. The metadata provides guidance in how to adapt the content for available resources.

2. SYSTEM DESCRIPTION

We begin by describing our system components in detail and examining the system functionality based on whether the content already exists and is stored in the content database or the content is created by the user with a mobile device, which are shown in Figures 1(a) and 1(b). The main parts of the system are the networked content server that we refer to as the content database, a networked display, and the user with a networked (wireless) mobile device. The main goal of the user is to view relevant information on an appropriate available display, whether mobile or fixed. We want to efficiently exploit the capabilities of the display in a resource constrained environment and render the appropriate image on the display based on metadata in the content database using a combination of image and video compression methods, and novel image rendering techniques. The resources that may constrain what we want to do include power (battery), computational bandwidth (on the mobile device and the display), display resolution, content resolution, and network bandwidth.



(a) (b)

Figure 1. Two user scenarios. (a) The content already exists and is stored in the database. (b) The content is acquired by the mobile device.

We are making the following assumptions in our system:

- **Content Database**: The database contains compressed images, video, and graphical models, and is connected to a high-speed network. The content is augmented by metadata that describes something about the content. For example, we know that a particular video contains a scene that was shot on the corner of Teal Avenue and Front Street in Lafayette, Indiana on May 10, 2002 at 4:57 pm. Our database will have a graphical model of that part of town that we can use to render a view of the scene. We are investigating what types of metadata we need to augment our content to adapt the content for the user. This project does not address hardware design for the content database (e.g., disk striping and robust storage). We will assume our metadata and scene descriptions will follow the concepts developed for MPEG-7 and MPEG-21.[7,23,30,32,38] Our initially deployed database uses JPEG and

JPEG2000 for image compression and MPEG-1, MPEG-2, and various versions of MPEG-4 (part 1, part 2, and part 10) for the video compression.

- **Displays**: Besides the display built into the mobile device, we have very high resolution displays available. These displays are attached to local networked computers that are used for rendering, caching, and acquiring information from the content database when appropriate.
- **Mobile Devices**: Our mobile device has a camera and is capable of acquiring both still images and videos and displaying them on a moderate to low resolution display. We also assume the device can determine its location either through GPS or other means and is connected to a high speed wireless network such as 3G or WiMAX.

Figure 1(a) shows the situation where the content already exists and is stored in the content database. The user may either download the content or stream the content to the mobile device. For example, the user may download a video of a news or sport event. This video may have associated metadata that provides some information as to when, where, or how the content was created. In a typical situation, the user would watch this video on the small screen on the mobile device. Our system would support the situation where the user may roam into an area with a large high resolution display where the content can be displayed. In a generic sense, one could think of this as a "display kiosk." For example, this display may be in a conference room, classroom or even in a public area. The user, with appropriate access permissions, would send the content (images + metadata) to the display and the display would then render the content. The display would use the metadata it received from the user and may obtain other metadata from the content database to construct the scene in a way that uses the full capabilities of the scene. For example, the video may be standard definition (SD) video (720x480 pixels per frame) but the display can display 1080i HDTV (1902x1080 pixels per frame). If graphical models exist for the scenes in the video, then these models can be used along the actual video to render the image in a way to use the full resolution of the screen. Alternatively, the mobile device may only have a 320x240 pixel display. We will use a similar approach to determine the best way to render the SD video on this display, rather than simply shrinking the image.

Figure 1(b) shows the situation where the content is acquired by the mobile device using the built-in camera. The user may want to display this content on the larger display. The user would connect to the display and send the content to the display along with metadata associated with the mobile device. This metadata would include the location of the device, date and time, and parameters of the device camera. The display would then acquire other metadata such as a scene model for the scene and then render the image.

3. ADAPTATION SYSTEM

Our system consists of a group of displays that are connected to the content database through a network. The displays actively determine how to present the information more effectively, with the guidance of the metadata associated with the content. This network may be wired or wireless. The displays are of different sizes and may have different network bandwidths. Additionally, some mobile devices are connected through a wireless network. The content may be acquired or generated by computers. The metadata is either created manually or extracted automatically. Our current system comprises two major components: content adaptation and resource-driven processing using the metadata.

3.1. Content Adaptation

3.1.1. Related Work

The concept of content adaption is not new. One of the first complete systems was described by IBM in 1999[27,28] adapts multimedia Web content to optimally match the resources and capabilities of various mobile client devices. The system had two main components, a data structure known as the InfoPyramid[27] and "customizer." The InfoPyramid provides a multimodal, multiresolution representation for the content items and their transcoded versions. The customizer selects the best versions of the content items from the InfoPyramids to meet the client resources. The content and adaption information was described using XML.

The system was very innovative, but from the point of view of rendering the content on the display, the images were re-scaled to match the display resolution.

Another approach based on 3D content adaptation was presented in.[18] This paper describes texture adaptation and streaming for three-dimensional applications using MPEG-4's Visual Texture Coding tool in conjunction with XML description techniques. Augmented features for content adaptation are supported, such as region selection, accompanied by resolution, and SNR settings. As a result, quality is optimized for the terminal's computing capabilities and display resolution (by either re-scaling the image or rendering with different qualities based on the compression scheme and QoS metrics in the network). Other approaches include content adaption based on the network adaptation to enhance ways of finding content in the network,[22] device independent models,[19] adaption for visually impaired users,[37] and transcoding and summarization tools for adapting content.[12] All of the above work concentrates on issues that involve network optimization and compression adaptation either through scalable compression methods or transcoding. The display adaptation is performed by re-scaling the image to fit the display. Our work involves adapting the displayed content through augmented graphical representations of the scene. In this paper, we will develop a system for resource constrained, user, task, and environment content adaptation.

Finally, content adaptation has been the focus of much of the recent MPEG-21 standard activities.[7,9,13,38] MPEG-21 aims at defining an open framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain. The MPEG-21 vision can thus be summarized as defining a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities.[7] The goal of MPEG-21 is thus: defining the technology needed to support users to exchange, access, consume, trade, and otherwise manipulate content in an efficient, transparent, and interoperable way. MPEG-21 does not specify tools or techniques for content adaptation but provides structures for describing a "language" for content adaptation based on XML.[9] In many cases, the adaption information is represented in terms of MPEG-7 descriptions.[6,30,32] MPEG-21 describes Digital Item Declarations (DID) and Digital Item Adaptation (DIA).

3.1.2. Content Adaptation with Metadata

Three issues must be addressed in our system: how to adapt the content based on the metadata, how are the metadata obtained, and how do we impose the resource constraints. We will first address the problem of how the metadata will be used and the other two issues will be discussed later.

For more effective communication of information, the presented content must be adapted based on the user, their task, and their location. However, this adaptation must also be driven by resource constraints: the content must be adjusted based on display resolution, network bandwidth, battery energy, etc. This is fundamentally different from changing the quality of service, such as suggested in.[15] A figure may show the scene when two children are playing with a ball. As more pixels are available, the contextual information may be added, for example, the playground. When enough pixels are available, both the focus and the context are shown. As the number of pixels decreases, the context is gradually removed, while the focus is kept.

Even though the above approach is fundamentally different from pixel-level manipulations, the adapted content is still based on the original images. We are investigating an approach for content adaptation using completely different representations, for example between images and sketches. It has been shown that simple sketches of human faces provide more accurate recognition than reduced and blurred photograph images of faces.[8] Our approach dynamically adjusts content by using different representations. We based this on our earlier study in illustrative visualization [20,21,33,34] to develop feature-based illustrative representations of images, graphics, and video for transmission and processing on displays of varying capabilities and different user tasks. By taking advantage of the human perceptual system's ability to reconstruct three-dimensional objects from a small set of edge cues, we can use feature sketches, stippling, and other illustrative processing to produce a compact, more effective, and more informative data representation than reduced/compressed images and video. We plan to explore point and line-based representations of complex volumetric datasets and polygonal geometric objects to provide a quickly rendered, compact representation of objects and volumes. Since points are the simplest rendering primitive and available in all graphics processors, they are a convenient and effective solution for developing adaptive rendering technology and have been recently used for remote visualization.[17] This work will build upon recent work in rendering for mobile devices[1, 10, 16] and, more specifically, in using lines and points for rendering on mobile devices[11, 14] to create a complete resource adapted illustrative, non-photorealistic rendering architecture and metadata description.

Our approach not only supports resource-driven location adaptation, but will also enable user and task adaptation, which are are crucial to effective information conveyance. We plan to expand upon our work in medical visualization adaptation based on user expertise and task[34] to general task and user adaptation rendering and image generation. Our initial work in this area generated different display representations for anatomy education (Figure 2(a)) and surgery simulation (Figure 2(b)) from the same medical scanner data, and allowed the level of complexity to vary for surgical simulation based on the expertise of the surgical resident, as can be seen in Figure 2.

Figure 2. Task adapted rendering of a foot dataset, showing an anatomical illustrations of the bones of the foot in (a) compared to an ankle surgery rendering (b). (c-e) Novice, journeyman, and expert temporal bone surgical simulation rendering adaptation.

We plan on retargeting video, graphics, and image sequences based not only on the resolution characteristics of the display, but also based on the user's task, device capability (battery, processing power, memory), network bandwidth, and the user's location. We are exploring methods for including illustration-based rendering techniques for compact, quickly rendered representations of geometry and larger-scale image primitives (e.g., brushstrokes). This process will first involve pre-processing the images, graphics, and video to generate appropriate metadata. We will also explore techniques for on-the-fly metadata generation and using heuristics for determining image importance based on simple image feature detection techniques. For memory-limited or bandwidth-limited devices, we can use this metadata and render from this representation on the local device. If power or processing capabilities are limiting factors, we can display unprocessed images, or importance enhanced images as appropriate. Using higher-level geometric or primitive-based metadata will also allow us to retarget images and graphics to larger displays than they were originally designed for by rendering more detail within the graphic, displaying contextual information, or related linked information.

To enable effective presentation of information that adapts to the display, processing capabilities, task, user preferences, and user location, we embed additional information, metadata, within the content. Metadata can include a variety of information, such as content descriptors, multi-level content description, importance information, and geometric information for computer generated images and animation. Such descriptions can be hierarchical for more flexibility in trading off detail and size of the metadata.

We use this metadata to enable us to take a novel approach to content adaptation. Traditional approaches to adapting images and video to varying display resolutions and capabilities usually just reduce or expand the source resolution to fit the display capability. Unfortunately, this approach is only applicable across displays with processing capability and display resolutions that only slightly differ (e.g., 1/2 the resolution). For instance, image averaging techniques over wider ranges of display resolutions produce images that are often too blurry or undetailed to be useful. Recently, Setlur et al.[31] have proposed a different image retargeting approach where they try to maintain image saliency as shown in their example image. In their approach, they average down the background to the target image size, they segment out the main focus objects in the image and resize them to fit within the image target size but recompose the scene to allow the foreground objects to be as large as possible. This is a promising alternative approach. Therefore, we plan to extend upon this recent work in image retargeting.

3.1.3. Metadata Generation and Extraction from Images and Video

In the previous section, we described how we would augment the graphical models with metadata and use the graphical information as metadata. We further extend this concept to the images and video in our content database and also the resource information. This metadata may include four types of information: (a) What are the expected resource requirements in terms of resolution, network bandwidth, computation capability, and energy consumption? (b) What is the emphasis of the content if the available resources do not reach the requirements? (c) Where can additional information be retrieved if the available resources exceed the requirement? (d) How can the content be regenerated with additional resources? In some cases, we may have to provide the metadata manually such as scene location and program transcripts. We are developing tools for automatic extraction of metadata. These tools are extensions of the work we did in developing the ViBE system.[36] Examples of the tools we are developing are:

- content description. As the amount of digital content increases rapidly, classification and organization are important and yet difficult. The descriptions may come from the original images or videos or other sources. One example is using a program schedule (a web site) to determine whether a program (video) belongs to sports, news, or drama.[24] Other examples are face detection and identification[2, 3, 35] and the integration of the video and audio tracks.[4, 5] We have developed tools that will also integrate this information with the video.[36] We intend to examine how these more "classical" approaches to content descriptions can be used in a content adaptation system for the display rendering.

- closed captions. We have developed an extensive set of tools for extracting descriptions of a scene from closed captioning information.[24–26] The tools use language models and information from the video frames.

- geometry and importance information. For computer generated images, video, and animation, geometric object and important information can be supplied to allow the generation of images at the appropriate resolution, rendering on the local device with power and display capabilities as factors, highlighting and rendering of appropriate information for the current user, task and location. For instance, for a cell phone display of a picture of a group of people, a sketch or caricature of the people can provide better recognition of the people in the photograph than a lossy compressed version of the image. Similarly, geometry and importance information can allow the rendering style detail to vary appropriately to help focus the user's attention and provide a better representation of the data than image scaling for screen resolutions varying from cell phones to wall-displays.

  To generate illustrative/sketch models of geometric models apriori, we will explore curvature and feature extraction techniques from multi-resolution geometric models, as well as a multi-resolution pre-rendered "depth peeling" approach to generate illustrative geometric primitives (points, lines) using multi-pass fragment programmability on PC graphics processing units.

  For image and video sequences, we can use feature extraction and geometry extraction techniques to generate higher-level geometric representations that can enable more compact representations, rendering at various levels of detail, combination with existing geometric models, and image-based rendering approaches. We also adapting and extending non-photorealistic rendering techniques for images and video to create illustrative point, line, and brushstroke representations of existing content within our database.

- resource requirements. The content can provide the information to determine whether simple image rescaling is appropriate or if more complex algorithms to refocus images or regenerate images are more suitable based on memory, battery, and processing requirements.

- user task profile. We will develop a metadata framework for specifying user task profiles to include task constraints, categories of relevant information, image rendering and scene profiles, and a hierarchy of appropriate augmented information to include based on available constraints.

3.2. Resource-Driven Processing

Existing studies on resource management assume little or no knowledge about the running applications. Hence, today's management schemes are mostly *reactive* in three steps. First, the system executes a program, downloads a file, displays an image, or plays a video. Second, the system observes the resource required for the content. Examples include a processor's utilization, the network bandwidth, or the amount of memory allocated. The third step is to respond by allowing or restricting the resources assigned for the content. If the processor's utilization is too low, the processor enters a lower frequency. If the network is too slow, a video is disrupted while waiting for frames being buffered in memory.

Our approach is unique because we *pro-actively* adjust the content for the available resources. When more resources are available, more content should be presented. The additional content can be the details, additional information, or the context of the information. One type of additional information is the resource requirements. We consider the following resources: display properties, computation speed, network bandwidth, storage, and battery energy. These resources are closely related. To use a higher resolution, a system needs more information, possibly downloading from the network at a high bandwidth. The system needs more computation capability to process the information. On a mobile device, the processing drains the battery and reduces the operational time of the system. To achieve better resource management, the information about resource requirements is embedded in the content. This section explains how to manage resources. Section 3.1 describes the changes that can be perceived by users.

The metadata stored in the content database provide the information that specifies the resource requirements. When content is transmitted to a smart display, the display communicates with the database to determine the resources available for the content. The database and the display have to "calibrate" the resources at run-time because it is not possible for the database to know every type of resource in advance. Moreover, the resources may change dynamically due to the environment such as noise in wireless channels. This calibration can be achieved by embedding small sections of calibration instructions in the metadata. The display can execute the instructions and transmit the results back to the database. The database then uses the results to select the most appropriate content to be displayed.

3.2.1. Display Properties

For a display, "resolution" is one of the most widely used properties to describe the hardware. We can describe an LCD monitor by its resolution as 1280×1024 pixels. In this paper, we generalize the concept of resolution from a hardware parameter to the number of pixels allocated to display content. This generalization extends our approach to dynamically select the right content. For example, consider a web page containing an image whose metadata allows content adaptation. When a user adjusts the size of a web browser, the image content is adapted with the new content. In this example, the hardware remains the same — still 1280×1024 pixels — but the number of pixels allocated to the image has changed. Our approach is different from existing HTML pages which directly specify the number of pixels used in a browser. Existing HTML pages have three drawbacks. First, if the browser is too small, the image spans beyond the browser and the user cannot see the complete image easily. The user has to scroll horizontally, vertically, or both. Second, if the actual resolution is higher, the additional pixels are wasted because the page cannot take advantages of the resources. Third, even if multiple images are used at different resolutions, they do not consider the fact that human beings perceive information differently for different sizes. For a low resolution, a sketch may convey information more clearly than a blurred image.

The second property of a display is the brightness. Many users dim the backlight of their mobile devices in order to prolong the battery lives. Because of this variation in brightness, the same content may appear very differently on different mobile devices. The second adaptation based on display properties is to adjust the content so that the information can be seen clearly. If the backlight brightness is reduced, the content's contrast should be enhanced. If the backlight is too bright (for example, when a mobile device is drawing AC power), the content's overall brightness may need to be reduced to prevent the content appearing "washed out." This adaptation can be performed by adjusting the content's color histograms or brightness histograms and this technique is well-known in image processing. Our approach emphasizes the importance of using metadata to guide such adaptation because some content cannot be arbitrarily adjusted based on the display's brightness, for example, medical images. Otherwise, the crucial information may be lost and the same content may produce different diagnoses when using different displays. Our system is unique because we use metadata to guide this adjustment. The metadata can specify whether adjustment is allowed and the constraints of the adjustment. A constraint can specify that the intensity histogram can only be shifted within a range and equalization is disallowed. Another constraint allows the histogram to be equalized but the standard deviation must be within 95% to 105% of the original content.

3.2.2. Network Bandwidth

The speed of a system is reflected in the delay and refresh rate. Most people have experience waiting for image downloading through a network. The current solution uses "progressive transmission": instead of sending the image one scanline at a time, the image can be sent in two phases. In the first phase, every other scanline is sent. The second phase sends the remaining lines, enabling users to see a degraded image (missing half of the lines) with only half the delay. Progressive transmission does not solve the problem of limited network bandwidth and it can only reduce the delay by at most 50%. In today's wide ranging network bandwidth, from Bluetooth 721 kbps to 1 Gbps wired local network, progressive transmission is insufficient for handling this wide range of differences. Hence, content adaptation is more attractive. Content adaptation for network bandwidth requires knowledge of the network. While the peak bandwidth is available from the type of the network used (such as Bluetooth, 802.11b, 802.11g, or 100 Mb Ethernet), the actual network properties can be affected by congestion or noise. Furthermore, computation speed (to be discussed later) also affects the delay and refresh rate. Two systems may have different refresh rates if one has a graphics accelerator even though the two systems have identical network interfaces.

We use metadata to instruct a system in determining what content to display. If the delay is too long or the refresh rate is too low, the content is adjusted to reduce the demand on the network. Similarly, if the network can support a higher demand, the content can be adjusted to take advantages of the network—a larger image, a high refresh rate, or more details. The metadata can also specify the constraints for displaying the content. In some applications, maintaining the refresh rate is more important than larger images. For example, decision making and medical diagnosis based on video sequences both require time-accurate presentation of information. To calibrate the available network bandwidth, the metadata can contain a small section of code requesting the smart display to respond. From the response, the content database and the display can determine the bandwidth and how to adapt the content accordingly.

3.2.3. Computation Speed

Content description (such as location of acquisition, geometry, objects, scenes, and rendering parameters) can be embedded with the content. The content can be displayed directly or regenerated before being displayed. This flexibility has many advantages. First, for some content, transmitting the descriptions and regenerating the content can significantly reduce the demand on the network. This is especially true as processors' speeds continue to grow while network bandwidths remain unchanged (particularly for wireless networks). Second, content descriptions may contain more details. The details allow the displays to adjust the presented content based on available resources. Third, some operations, such as 3-dimensional manipulations, are possible only if the content descriptions are available.

The computation speed and capabilities are crucial factors in deciding whether to regenerate content from its description or display the content directly. Content should be regenerated only if the computation speed can meet certain performance criteria (delay and refresh rate). Computation capability spans a wide range from low-power processors in cellular phones to multiple GHz superscalar dual-core desktops with accelerated multipipe graphics processors (e.g., PC graphics cards). It is not possible to customize content for each processor. Even for the same processor, the performance may vary due to the network connection, noise in the wireless channel, or congestion. Therefore, our approach uses a small segment of the content and its description to calibrate the computation speed at run-time. Before displaying the content, the system executes this small segment to determine the performance that can be achieved in this system. With this information, the system can determine whether to display the content directly or to regenerate it. Our approach also facilitates energy conservation for mobile devices. When the computation capability exceeds the needed performance, the mobile devices can reduce the power consumption by lowering the processors' clock frequencies and voltages without degrading the performance.

3.2.4. Storage and Battery Energy

Storage is another important factor in determining content adaptation. First, the metadata require additional storage. Second, storage is needed to keep the intermediate results for content adaptation. Third, as we have shown in our study,[29] storage has a direct impact on quality of service of videos and energy savings. When more storage space is available, more future frames can be buffered so that the frame rate can remain steady even though the network bandwidth fluctuates. If the storage space is insufficient, it can be more acceptable to watch the video with a lower resolution, instead of a reduced frame rate. For mobile devices, battery energy is a precious resource so it is an important factor for content adaptation. When the battery energy is considered, we can select several different ways to adapt the content. Two simple approaches are reducing the frame rate or the resolution. As explained earlier, these two solutions may be unacceptable, depending on the types of content. Our system also includes changing the representation, for example, 3-dimensional rendering to sketches. The metadata can also specify whether it is better to maintain the quality without finishing the video, or finishing the video at a lower quality.

4. CONCLUSION

We feel that content adaptation is one of the last barriers that needs to be addressed for "true" collaboration in heterogeneous environments. The concept of "anytime" at "anywhere" with "any device" requires that the content be adapted to meet the needs of the system and the user. In this paper we have described the system we are developing at Purdue University to address the types of problems presented by content adaptation.

REFERENCES

1. T. Akenine-Möller and J. Ström. Graphics for the Masses: A Hardware Rasterization Architecture for Mobile Phones. *ACM Transactions on Graphics*, 22(3):801–808, 2003.
2. A. Albiol, C. A. Bouman, and E. J. Delp. Face Detection for Pseudo-Semantic Labeling in Video Database. In *International Conference on Image Processing*, pages 607–611, 1999.
3. A. Albiol, L. Torres, C. A. Bouman, and E. J. Delp. A Simple and Efficient Face Detection Algorithm for Video Database Applications. In *International Conference on Image Processing*, pages 239–242, 2000.
4. A. Albiol, L. Torres, and E. J. Delp. Combining Audio and Video for Video Sequence Indexing Applications. In *IEEE International Conference on Multimedia and Expo*, pages 353–356, 2002.
5. A. Albiol, L. Torres, and E. J. Delp. The Indexing of Persons in News Sequences using Audio-Visual Data. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 137–140, 2003.
6. O. Avaro and P. Salembier. Mpeg-7 systems: overview. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):760–764, June 2001.
7. J. Bormans, J. Gelissen, and A. Perkins. MPEG-21: The 21st century multimedia framework. *IEEE Signal Processing Magazine*, 20(2):53–62, March 2003.
8. Bruce Gooch and Erik Reinhard and Amy Gooch. Human Facial Illustrations: Creation and Psychophysical Evaluation. *ACM Transactions on Graphics*, 23(1):27, 1 2004.
9. I. Burnett, R. V. de Walle, K. Hill, J. Bormans, and F. Pereira. MPEG-21: goals and achievements. *IEEE Multimedia*, 10(4):60–70, October-December 2003.
10. D. Hekmatzada and J. Meseth and R. Klein. Non-Photorealistic Rendering of Complex 3D Models on Mobile Devices. In *Annual Conference of the International Association for Mathematical Geology*, pages 93–98, 2002.
11. J. Diepstraten, M. Groeke, and T. Ertl. Remote Line Rendering for Mobile Devices. In *Computer Graphics International*, pages 454–461, 2004.

12. A. Divakaran, A. Vetro, and T. Kan. Towards maximizing the end-user experience. In *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo*, volume 3, pages 2147 – 2150, June 2004.
13. S. Dobravec, B. Marusic, J. F. Tasica, A. Jere, M. Meza, and T. Finkst. The MUFFINS online store - buy it in the MPEG-21 way. In *Proceedings of the 12th IEEE Mediterranean Electrotechnical Conference*, volume 2, pages 701–704, May 2004.
14. F. Duguet and G. Drettakis. Flexible Point-Based Rendering on Mobile Devices. *IEEE Computer Graphics and Applications*, 24(4):57–63, 2004.
15. J. Flinn and M. Satyanarayanan. Energy-Aware Adaptation for Mobile Applications. In *ACM Symposium on Operating Systems Principles*, pages 48–63, 1999.
16. G. Hesina and D. Schmalstieg. A Network Architecture for Remote Rendering. In *International Workshop on Distributed Interactive Simulation and Real-Time Applications*, pages 88–91, 1998.
17. G. Ji, H.-W. Shen, and J. Gao. Interactive Exploration of Remote Isosurfaces with Point-based Non-Photorealistic Rendering. In *Joint Eurographics - IEEE TCVG Symposium on Visualization*, 2003.
18. G. Lafruit, E. Delfosse, R. Osorio, W. van Raemdonck, V. Ferentinos, and J. Bormans. View-dependent, scalable texture streaming in 3-D QoS with MPEG-4 visual texture coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(7):1021 – 1031, July 2004.
19. T. Lemlouma and N. Layaïda. Context-aware adaptation for mobile devices. In *Proceedings of the 2004 IEEE International Conference on Mobile Data Management*, 2004.
20. A. Lu, C. J. Morris, D. S. Ebert, P. Rheingans, and C. Hansen. Non-photorealistic volume rendering using stippling techniques. In *Proceedings of the conference on Visualization '02*, pages 211–218. IEEE Press, 2002.
21. A. Lu, C. J. Morris, J. Taylor, D. S. Ebert, C. Hansen, P. Rheingans, and M. Hartner. Illustrative interactive stipple rendering. *IEEE Transactions on Visualization and Computer Graphics*, 9(2):127–138, Apr-Jun 2003.
22. M. Lyijynen, T. Koskinen, S. Lehtonen, and J. Pesola. Content adaptation on LANE active network platform. In *Proceedings of the 7th International Conference on Telecommunications*, volume 2, pages 721 – 724, June 2003.
23. B. S. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7*. Wiley, 2002.
24. A. Martone and E. Delp. An overview of the use of closed caption information for video indexing and searching. In *Proceedings of the CBMI Fourth International Workshop on Content-Based Multimedia Indexing*, Riga, Latvia, June 2005.
25. A. Martone, C. Taskiran, and E. Delp. Automated closed-captioning using text alignment. In *Proceedings of the SPIE International Conference On Storage and Retrieval Methods and Applications for Multimedia*, pages 108–116, San Jose, CA, January 2004.
26. A. Martone, C. Taskiran, and E. Delp. Multimodal approach for speaker identification in news programs. In *Proceedings of the SPIE International Conference On Storage and Retrieval Methods and Applications for Multimedia*, San Jose, CA, January 2005.
27. R. Mohan, J. R. Smith, and C. H. Li. Adapting multimedia internet content for universal access. *IEEE Transactions on Multimedia*, 1(1):104–114, March 1999.
28. R. Mohan, J. R. Smith, and C.-S. Li. Content adaptation framework: Bringing the internet to information appliances. In *Proceedings of the 1999 Global Telecommunications Conference*, December 1999.
29. N. Pettis, L. Cai, and Y.-H. Lu. Dynamic Power Management for Streaming Data. In *International Symposium on Low Power Electronics and Design*, pages 62–65, 2004.
30. P. Salembier and J. R. Smith. Mpeg-7 multimedia description schemes. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):748–759, June 2001.
31. V. Setlur, S. Takagi, R. Raskar, M. Gleicher, and B. Gooch. Automatic Image Retargeting. In *Conference Abstracts and Applications of ACM SIGGRAPH*, 2004.

32. T. Sikora. The mpeg-7 visual standard for content description: an overview. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):696–702, June 2001.
33. N. Svakhine and D. Ebert. Interactive volume illustration and feature halos. *Pacific Graphics '03 Proceedings*, 15(3):67–76, 2003.
34. N. Svakhine, D. S. Ebert, and D. Stredney. Illustration motifs for effective medical volume illustration. *To appear in IEEE Computer Graphics and Applications*, 2005.
35. C. Taskiran, A. Albiol, L. Torres, and E. J. Delp. Detection of Unique People in News Programs Using Multimodal Shot Clustering. In *International Conference on Image Processing*, pages 697–700, 2004.
36. C. Taskiran, J. Chen, A. Albiol, L. Torres, C. A. Bouman, and E. J. Delp. Vibe: a compressed video database structured for active browsing and search. *IEEE Transactions on Multimedia*, 6(1):103–118, February 2004.
37. T. C. Thang and Y. M. Ro. Visual content adaptation for flow vision users in MPEG-21 framework. In *Proceedings of the IEEE International Conference on Image Processing 2004*, volume 2, pages 993–996, October 2004.
38. A. Vetro. MPEG-21 digital item adaptation: enabling universal multimedia access. *IEEE Multimedia*, 11(1):84–87, January-March 2004.