INTERACTION WITH CONTENT-AUGMENTED VIDEO VIA OFF-SCREEN HYPERLINKS FOR DIRECT INFORMATION RETRIEVAL

Dirk Balfanz, Matthias Finke

Department of Mobile Information Visualization
Computer Graphics Center, Rundeturmstrasse 6
64283 Darmstadt, Germany
{Dirk Balfanz, Matthias.Finke}@zgdv.de http://www.zgdv.de/miv

ABSTRACT

Video presentation systems currently available provide users with only a few functionality options. With the introduction of an Internet-based back channel, the way is now open for wide-spread interactive video systems. The convergence of the Internet with broadcast services gives rise to a variety of new application types. It extends the opportunity for a client to retrieve additional video content-related information or to affect the content of a video presentation itself, whether it is a live broadcast or a video-on-demand. The system presented in this paper represents such an interactive application with a number of new options concerning local and global functionalities. To influence video content, a "off-screen" hyperlink technique is introduced. This technique enables the annotation of hyperlinks on a separate display from the actual information with which it is linked. In the following, a scenario is given demonstrating the advantages of such a technique.

Keywords: adaptive additional information, personalized presentation, bi-directional remote control, dynamic display, live content manipulation, "off-screen" hyperlinks, Interactive video / TV

1. INTRODUCTION

Such new communication channels as the Internet or mobile networks, along with related new technologies, have increased the amount of accessible information and its availability to an extent never known before. Although some effects of the so-called Information Society, such as - for instance - the information flood, are perceived to be negative, the same progress has developed new interactive and multimedia ways of working with this virtual information space.

Information retrieval via the Internet and interactive information representation in multimedia applications, e.g. in multimedia encyclopaedias, are well-known and widely used. The new media found approaches to more user-centered and interactive handling of information.

On the other hand, the "old media", such as classic TV, has developed, as well. Cable networks and satellite broadcasting multiplied the available channels. However, in contrast to the new media, the interaction paradigm in TV broadcasting has stayed the same.

Many attempts to introduce Interactive TV failed due to fact that today's users are already accustomed to handling information in a more personalized and interactive way. Past approaches were chiefly limited to Video-on-Demand (VoD) functionality, Home Shopping applications or Electronic Program Guides (EPGs). This range of interactivity does not reach the level that other media are already presenting. For future applications in the area of Interactive video / Interactive TV, it is crucial to provide interaction with the content and not only with the media itself.

From this perspective, Interactive video / TV should be understood as the interaction of the user with the current video content through application features.

Furthermore, it is possible to distinguish between local and global interaction.

Local interaction contains all necessary information and processing components; one example could be interactive CD-ROM applications. Global interaction is used by distributed applications, which involve

client and server components, as well as a bi-directional communication channel.
In principle, interactive TV is open to support both levels of interaction.

Only a few years ago, there was no way for any system architecture to enhance typical broadcast scenarios, in particular with global interactivity, without introducing new infrastructures.
In general, today's communication channels are still used to carry primarily specific services (e.g. fixed bandwidth channels like satellite for transmission of fixed bandwidth signals as TV, videoconferencing or phone). The more or less fixed relationship between provider, channel and consumer can be described as follows [1]:

- selected channels are chosen for the whole transmission; interaction is only supported on dedicated channels,
- the more users are supported by a system, the less interactivity is provided,
- there is a clean cut between unicast, multicast and broadcast on different channels,
- there is a fixed relation between information source and information consumer.

The internet is now blurring these distinctions more and more and is creating new opportunities to integrate different channels and services. Approaches can be classified into two groups [2]:

- use of one transmission channel and one or more hardware devices for several services;
- use of several transmission channels and one single hardware device for the presentation of several services.

In the following section, an approach to enable interaction with video content is presented, based on dynamic provision of content-related hyperlinks. It is able to provide global interaction, based on a system architecture that uses broadcast services and the Internet in an integrated way.

2. RELATED WORK

Videotext was a technique developed in the early 80's [3] to provide access to modified TV sets or to remote computer databases from computer terminals. Today this form of information retrieval upon user request is still widespread in Europe. It provides interactivity while watching TV content. Due to the missing back channel, only local interaction functionality is provided. Another disadvantage is its limited capability to display information content, since videotext systems provide only text, color and crude block graphics. The system presented here offers global interactivity by introducing a back channel. In addition, the representation of additional information is strongly improved and can be adapted to user preferences.

From 1995 to 1997, the Microsoft Interactive TV System (MITV) [4] was deployed in a trial partly in the USA and Japan. An ATM network was integrated to meet the needs of distributing video content, while simultaneously introducing a back channel for every client. Set-top boxes were used in the trial of the MITV system. The applications running on this system were:

- Electronic Program Guides (EPG);
- Movies On Demand (MOD) applications;
- a Web browser was developed, but never employed for the trial.

The system presented in this paper integrates settop boxes and back channels too. The main difference is that the additional information within a video presentation is fully controlled by the user. In other words, the video presentation consists of video content and additional information, which are dynamically assembled, individually according to every single client.

A prototype application that combines a PDA, a settop box and a television has been developed for multimedia presentation by Robertson [5]. Here the television was only used to present images and video content. The actual selection and presentation of additional information content was done on a PDA. Furthermore, the application could run without the deployment of the television and the settop box. A disadvantage here is the poor capability of presenting the additional information only on a PDA's display with regard to its size and resolution.

3. SCENARIOS

To depict the viability of interactive TV, some possible application scenarios are presented in the following section. The functional scope of these scenarios is only partly covered by the described architecture of the next paragraph. Other parts are currently under research and will be referred to in the outlook section.
The main idea, garnered from the understanding of interactive TV as presented in the introduction, is to allow the former TV consumer to become an information user and to enable him to play an active part in the video presentation.
Some possible areas of this interactive broadcast format could be (business) information, education, edutainment or sports.

- *Advance Business TV*
  Current formats like Bloomberg TV, for instance, are presenting a lot of different information in parallel in the multi-window style today's users are accustomed to from PC-applications. The next logical step in interactive TV would be to provide the facility of choice. The user could be presented only that information he wants to see. The interaction could be managed via a two-way remote control presenting the current choices, or future choices, as well, by "on-screen-clicking" via special remote devices.

- *Edutainment*
  Nature or other popular scientific reports could be enriched with more detailed information (text or graphics) on specific topics. There could also be references to further media (videos, books, CD-ROMs, etc.) along with the presented topic that the user can purchase by using some kind of home-shopping functionality. This additional content would be visualized via OSD (on-screen-display).
  As in the first example, the user could interactively choose those parts of additional information he is most interested in via his remote control or visual on-screen hyper-links.
  The same concept would be valid for tele-teaching. Users would have the choice of getting a more personalized learning content. The upward channel could also be used for online changes to the course of a lesson. If, for instance, most viewers are concentrating on some specific topics by selecting the additional reference information, the system could choose the more appropriate of several parallel plot-threads analogue to multimedia applications.

- *Sports*
  The user might get features, such as clicking on players in a football game and receiving detailed information about the selected person, graphics showing aspects like the hit-rate, last year's results of the team playing or statistics like percentage of ball ownership during the current match. The user might also be provided with links to offers for purchasing fan articles or other media about the team or a certain player.
  More sophisticated systems could produce dynamic information like annotating the speed on specific cars in a race or enhancing the video of a 100m run with a virtual runner shown performing the last world record.

These scenarios provide an impression of the benefits for both TV consumers and the suppliers or broadcasting corporations. Furthermore, on the more technical level, one can derive the aspects, which have to be addressed when referring to the system architecture:

- transmission of the additional information downstream to the user and upstream to the broadcaster;
- presentation type of the additional information (on-screen, "off-screen");
- interaction type (e.g. choosing from a list vs. point-and-click);
- enabled functionality (e.g. presenting information related to current video content or more general information, changing the look of the video-content, changing the plot of video content).

4. SYSTEM ARCHITECTURE

As mentioned in the introduction, the system presented in this paper combines the Internet with broadcast services in order to form a new kind of interactive application platform. The figure below gives a general overview of the system prototype.

General System Overview
Figure 1

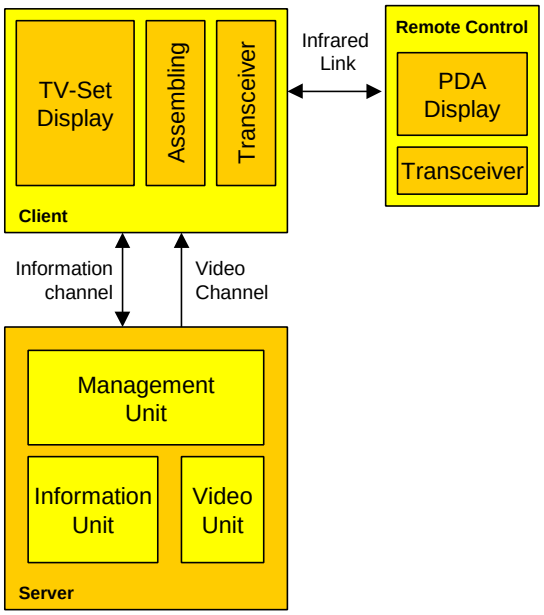
The main idea is to supply the user with new functionalities in order to access specific additional information that enriches the video content -- or to give him a certain level of control over the course of the video presentation, regardless of whether it is a live source or video-on-demand. These new functionalities have the main task to encourage, but not to force, the user to assume an active role within a video presentation.

In contrast to classic TV or video-on-demand systems, the described architecture allows a change in the video presentation itself, either simply in appearance or in content, as well.
Considering video presentation today, we can separate the actual video content from the information part that is given as an addition [6]. Conversely, a video presentation void of additional information would be more or less useless in many cases. Think about a tennis match where the information about the current standing is missing. Or, think about a basketball match where the time left to play will never be shown. This presentation would never be as interesting as it is with the information supplied today. The presented information itself can be divided into two main groups:

- *Time-dependent information*: This information type is of dynamic character. It is information, which is often directly derived from the current video content. It is not predictable in any sense and always time-dependent.
- *Overall information*: This information type is related to the video content, as well, but it is not extracted from it. Mainly, it has static character and will not change during the course of the presentation. Due to its static character, this information is usually generated before the presentation and can be reused for other presentations dealing with a similar topic.

The system presented here provides the user with the option of taking advantage of additional information as an interactive part of the video content. This means he or she determines how and when information is presented while watching video content.
The challenge was to find a solution that would provide a user with a simple yet effective way to access the information resources upon his request or -- in other words -- to "direct his own video presentation" [7].

The figure below shows the client - server architecture, which will be described in more detail in the following section.



System Overview
Figure 2

The main components are the server unit, which may be connected to live sources and/or video-on-demand systems, an advanced TV set [8] capable of Internet access, and a bidirectional remote control including a dynamic animated screen, displaying the off-screen hyperlinks.
The server side is in charge of delivering the video content and the corresponding additional information. A primary task for the server is to handle the synchronization between these two information flows. The server architecture is designed as a distributed system, which means that the actual video server unit might be located at a different location from the information server.
A video channel is established between the server side and all clients. This channel carries content, which is the same for all clients accessing it. So far, this channel is based on multicast streaming technologies [9] using Internet connections. In the future, the video content could even be carried by broadcast system via satellites.

A second channel defines the point-to-point connection between each client and the server. The channel realizes the data exchange underlying a point-to-point connection, since it carries the information that is determined for a single client and not for a group. This point-to-point communication is based on Internet connections.
The client side is equipped with an advanced TV-set capable of accessing the Internet. Advanced TV-sets from a large number of producers can already be found on the market. For instance, the German company Loewe [10] has an advanced TV-set in their product line not only equipped for Internet access, but also capable of running Java Applets.

Another interesting platform worth mentioning in this context is the multimedia home platform MHP [11]. The advanced TV-set of Loewe was chosen for the prototype solution of the system. Besides the two channels towards the server, the advanced TV set is serving a bi-directional infrared link, which establishes a connection to a remote control.
As a remote control device within this system, a common PDA, the Palm [12], was chosen. This device type has some advantages, for instance, a built-in, bi-directional infrared transceiver and a rather large touch screen for its size. In addition to its already great popularity as a consumer product, the Palm provides a strong developer environment including a large number of utility tools.
One of the PDA display tasks within this application is to show the so-called *charts*. A chart is a graphical surface containing icons and / or text elements. An icon or a text represents an "off-screen" hyperlink to additional information [13]. This means with a click on these hyperlinks within the display area, a user requests additional information from the system, like a textual hyperlink within a web page. As a second option for selecting these hyperlinks, the PDA buttons below the display can be used, as well. Usually the charts are divided into thematic subject groups providing a more structured overview about the possible information content accessible within the video presentation. The figure below shows a typical chart on a PDA device.

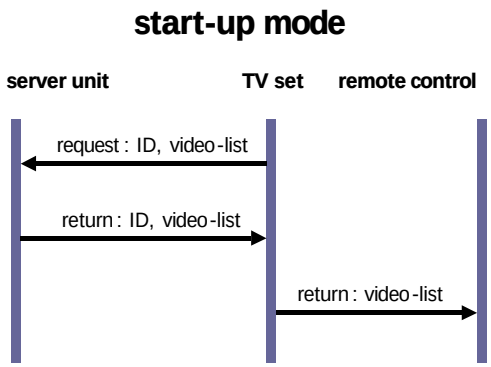
Typical Chart
Figure 3

In addition to displaying charts, a PDA screen can also show additional information. The intention here is only to show small portions of information content. The TV screen primarily presents all requested information which can be of different media formats like audio, video, image, graphics or even 3D animations.
With the PDA, a user requests not only additional information, but also defines his or her personal system preferences. Personal system preference capability is a great advantage for this application, since different user have different interests and background knowledge about the same subjects. When users retrieve identical information linked to the video content, it is always a compromise for the audience, not an ideal solution. Consequently, adapting additional information according to individual user needs results in more satisfaction. The system preferences that can be set up by the user according to his needs can be divided into three logical main groups [14]:

- *User preferences*: A user might change the language in which the information is displayed or choose certain topic categories of interest. The latter option would lead to adaptive information content according to the personal interests of a user. Other options might include the display duration of the information content, etc.
- *Remote Control preferences*: These preferences define the remote control itself, for instance, color or grey scale displays, the display size or the assignment of the buttons which might be useful for sinistrals.
- *TV preferences*: The preferences of the TV set can be compared with the ones of the remote control. Even here certain settings in the preference list allow the definition of the screen size and the TV aspect ratio, for instance.

The process flow of the system is built upon three modes, the *start-up mode*, the *initialize mode*, and the *process mode*. In the following, we will describe how the system behaves within these three modes.

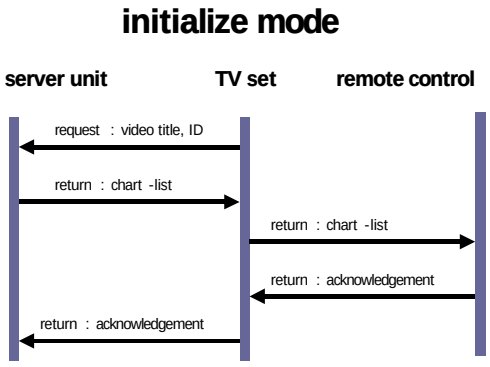
The *start-up mode* will occur when a user switches on the TV set for the first time. The client announces itself at the server's management unit by requesting an ID and a *video list* over the information channel. The management unit generates a unique ID. This user ID identifies the client side in the system and has a live-time that ends when switching off the TV set. Further more, the video unit generates the video list with the current content that can be presented. Both, the video list and the unique user ID are sent in return to the request to the client side.



Startup Mode
Figure 4

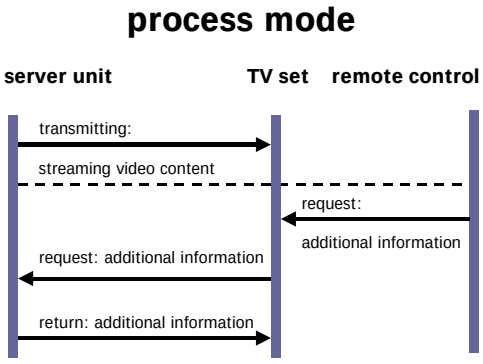
Today's broadcast system differs from packet switch networks, like the Internet, in which the content is transmitted once and can be received by every client without requesting it.

The *initialize mode* presented here is based on a packet-orientated network concerning the video content; consequently a user request is necessary. With the integration of a broadcast network within this project, the initialize mode would change. Every time a user requests a new video presentation, which can be a live source or video on demand, by looking up the video list, the system turns into initialize mode.
This user request contains the user ID and a name, identifying the content source. On the server side, the management unit receives the request and contacts both units, the information unit and the video unit. Before the video content is transmitted, the charts have to be downloaded on the client side's remote control. The information unit keeps a so-called *chart-list* for each video content. The chart-list contains all charts used as a container for the hyperlinks (icons and text elements) in order to retrieve additional information related to the video content.
This list is sent to the management unit, which transmits it over the information channel to the client side. The advanced TV set on the client side receives the chart-list and passes it via the infrared link to the PDA. The charts containing the hyperlinks are then shown on the PDA's display. When the PDA device is set up and ready for further processing, the management unit receives an acknowledgement. Now, the video presentation is ready to start and the system changes into process mode.



Initialize Mode
Figure 5

Process mode starts when the server side transmits the video content towards the client side. At this stage of the project, we are streaming all video content via the Internet using multicast technologies. When receiving the video content on the client side, the user is able to request information. With a click on the hyperlinks within each chart, a *user request* is generated and sent to the TV set, which passes it to the server side. After receiving the request, the management unit informs the information unit. This unit has to distinguish whether it is *time-dependent information* or *overall information*. Overall information can be sent back directly to the management unit, whereas time-dependent information must be generated in advance. When this information is available the management unit transmits it towards the advanced TV set.
The information is presented primarily on the TV screen with the help of pixel manipulation techniques in real-time using a Java application supported by the *Java Media Framework*.



Process Mode
Figure 6

The Figure below shows the interplay of the remote control and the advanced TV set. The displayed information is usually of transparent type as shown in the figure.

Interplay TV and Remote Control
Figure 7

5. FUTURE WORK

In the future, the project will partly focus on the objects within the video scene. So far, additional information is linked to the video content, but not linked to a video object in particular. The MovieGoer project [15] presents some basic work on this subject.
When linking objects in a video scene with additional information, one has to think about the annotation of these hyperlinks. Since video objects might move within a scene, the annotation will not be as simple as it is for textual hyperlinks.
From the hardware point of view, the Multimedia Home Platform MHP, currently under development, is an interesting application. In the future, this platform will be evaluated for the deployment of our system.
For multimedia applications, the standard MPEG-4 [16] establishes new options for an interactive video environment and compression techniques. Consequently, the analysis of this standard is one of the major aspects in the further development of the project.

6. CONCLUSION

We have presented a system that supports interactive video presentations. All necessary system components are available on the market today. Due to the use of both networks, the Internet and broadcast services, an ideal solution was presented for the combination of unicast and multicast channel transmission referring to the the video content and the additional information. Furthermore, the packet-orientated network, the Internet, is used to establish the back channel for data exchange. The bandwidth requirements for the applications concerning the two networks in use are fulfilled.
The introduction of a new kind of hyperlink, the *off-screen* hyperlink, has been presented and its advantages were demonstrated in detail. In addition, scenarios were given in which off-screen hyperlinks have been integrated. So far, a PDA is used serving a infrared link towards an advanced TV set capable of Internet access. The off-screen hyperlinks are displays on the PDA's screen. The actual requested information is usually presented on the TV screen.
Considering the convergence between the Internet and Broadcast services, the presented system really merges both network types in order to provide a video presentation combining video content with direct information retrieval upon user's request by implementing off-screen hyperlinks. Furthermore, time-dependent information is generated in real time out of the video content. This type of information, along with overall information, is directly related to the video content.

7. AUTHORS

Dirk Balfanz is the head of the department Mobile Information Visualization at the Computer Graphics Center in Darmstadt, Germany. In 1992, he graduated with a degree in Electrical Engineering from the University of Dortmund. After developing cordless DECT applications at Bosch Telecom, he worked as a researcher in the area of Geographic Information Systems at the Fraunhofer Institute for Computer Graphics. His research interests include use and extraction of (geo-) graphic meta data and context-driven visualization.

Matthias Finke graduated with a Master's degree from Chalmers University in Gothenburg, Sweden. He worked at Ericsson Mobile Data in the research department for Internet applications, then became a research assistant at the Computer Graphics Center in Germany in 1999. Matthias Finke is involved in the field *of Internet Interactive Broadcasting*. He is currently engaged in research concerning interactive digital video environments for Internet broadcasting applications.

REFERENCES

[1] Gerfelder, N.: Video Applications in the Era of Computer Networks, Proceedings 137th SMPTE Technical Conference and World Media Expo, Sep. 1995

[2] Gerfelder, N.: Video Integration and the Use of Various Delivery Channels, Proceedings of the European SMPTE Conference, Cologne, Germany, 1996

[3] ZDF, Geschichte des ZDF, ZDF Service, Mainz, Germany, http://www.zdf.de, 2000

[4] Jones, M.: The Microsoft Interactive TV System: An Experience Report (1997), Microsoft Research, Microsoft Corporation, Redmond, July, 1997

[5] Robertson, S., Wharton, C., Ashworth, C., Franzke M., *Dual Device User Interface Design: PDAs and Interactive Television*, Conference on Human Factors in Computing Systems, April, Vancouver, Canada, 1996

[6] Höhl, A.: "WWW Application for Interactive Video", Diploma Thesis, TH-Darmstadt, 1996

[7] Sawhney, N., Balcom, D., and Smith, I.: HyperCafe: narrative and aesthetic properties of hypervideo. The seventh ACM conference on HYPERTEXT, Proceedings, Bethesda, MD USA, 1996

[8] O'Driscol, G.: The Essential Guide to Digital Set-top Boxes and Interactive TV, Prentice Hall PTP, USA, 2000

[9] Hefta-Gaub, B.: The Real Media Platform Architectural Overview. Real Networks Conference, San Francisco, USA, 1997

[10] Loewe Opta GmbH, Kronach, Germany, http://www.loewe.de, 2001

[11] DVB, "Multimedia Home Platform (MHP)", TS 101 812 (MHP), http://www.dvb.org/standards/index.html, 2001

[12] Palm, Inc. Corporate Headquarters Santa Clara, CA USA, http://www.palm.com, 2001

[13] Bove, Jr., V.M., Dakss, J., Agamanolis, S., and Chalom, E.: "Adding Hyperlinks to Digital Television", SMPTE Journal, pp. 795-801, Nov. 1999

[14] Freytag, C., and L. Neumann: "Resource adaptive WWW access for mobile applications", Computers & Graphics, Vol. 23 (1999) pp. 841-848

[15] Finke, M.: An Architecture of a Personalised Dynamic Interactive Video System, Digital Content Creation Conference, Bradford, Britain, 2000

[16] Cosmas, et.al., „CUSTOMTV With MPEG-4 and MPEG-7“, Colloquium on interactive television, IEE, Savoy-Place, London, UK, 1999