VIZARD-EXPLORER: A tool for visualization, structuring and management of multimedia data

Herwig Rehatschek, Gert Kienast, Alfred Wertner

E-mail: {Herwig.Rehatschek, Gert.Kienast, Alfred.Wertner}@joanneum.at
WWW: http://iis.joanneum.at
Institute of Information Systems & Information Management
JOANNEUM RESEARCH, Steyrergasse 17, A-8010 Graz, Austria

ABSTRACT

The price for digital cameras (for still images as well as for videos) has significantly dropped over the period of the last few years. In parallel tremendous advances have been made in the standardization of storage formats and high-compression codecs for digital images and digital video. The demand for organizing, managing, searching and visualizing within multimedia content (audio, video, images but maybe also documents related to images, videos) is automatically generated with growing amounts of produced data. We propose a tool - the VExplorer, respectively - which offers functionalities strongly related with the management and organization of multimedia content. Features of the VExplorer include new navigation concepts within digital video essence (O-M images), an interface for creating sketches, a search interface (content based retrieval & text) and a visual browsing tool for large archives.

KEY WORDS

Digital video, visualization, navigation, browsing, story sketching,

1.   INTRODUCTION

Since the development of digital video technologies in the 1980s the price for digital cameras (for still images as well as for videos) has significantly dropped over the period of the last few years. In parallel tremendous advances have been made in the development and standardization of storage formats and high-compression codecs for multimedia data, specifically for digital images and digital video. With the availability of these formats and codecs very efficient image and video compression techniques were introduced (MPEG-1/2/4 [3], DV, Wavelet, DivX, Cinepak, Indeo, etc.) allowing compression rates up to 1:50 or even more could be achieved [1]. Together with the decrease of HW prices in the computer sector this opened the door for a very large community of home users to produce digital videos and images. After a certain time of usage (the amount depends on how extensive the digital devices are used) the demand for organizing, managing, searching and visualizing within multimedia content (audio, video, images but maybe also documents related to images, videos) is automatically generated. The R&D project VIZARD targets exactly this requirement with a strong focus on fulfilling the requirements of home users and semi-professional users. With "semi-professional" all users are addressed who have a need to organize and maintain their content but are not professional video editors and post-producers. Typical semi-professional users are journalists, reporters, researchers and archivists.

The proposed VIZARD-EXPLORER (VExplorer) will be a comprehensive "video collection search, organization and management" tool for organizing local content and in order to pre-structure material in a repository. A key focus is laid to provide a user with logical views on the local archive on any kind of data (e.g. video, audio, text, images, programs). These views can be customized as hierarchically structureable collections representing temporal orders. Next to this the VExplorer enables the user to draft a story or a sketch by selecting and structuring material. During this phase the user can take advantage of efficient search mechanisms provided by the tool. Searching for content should be possible either by similarity (query by image) or by searching annotations created with a special annotation tool - the VAnnotator which is also part of the VIZARD R&D project, but is not covered by this paper. The entire repository containing the stories can be persistently stored in XML format. An export function supporting the internationally standardized MPEG-7 format is under development.

The paper is organized as follows: In chapter 2 an overview on related work is given. It is also clearly explained how our work differs from this work. In chapter 3 the results achieved so far in the VIZARD R&D project concerning the VExplorer component are presented. Chapter 4 gives an outlook on future work, chapter 5 summarizes important observations and conclusions.

2.   RELATED WORK

A system which allows a navigation within video essence is the Hyper-G system [4] which was initially developed at the Graz University of Technology and is now commercially sold as Hyperwave [5]. The Hyper-G system supports the definition of hyperlinks as known from the World Wide Web within digital videos in order to reference to other information stored on the server, e.g.. additional text information on the current scene.

In connection with the visual browsing tool of the VExplorer described in chapter 3.5 a project currently being executed at the University of Arizona [10] with the name "Vizard" is of interest (Note: despite the project has the same name as our R&D project and scientific relevance to our work the two projects are not administratively connected in any sense). The researchers propose the methods of multidimensional scaling and hierarchical clustering in order to represent visually multidimensional information. However, this is a general purpose approach to manage information. A specialized approach for multimedia data does not exist so far.

The VideoQ system offers a visual search which is not limited on a simple query by example but allows the user to formulate 'sketches' comprised of the objects of interest [2]. Despite the system is very innovative in the sense, that all annotations necessary for the search are generated automatically it is limited to a search of simple objects. It fails when too many objects (e.g. a group of people) are displayed in the frames. Also – in comparison to the VExplorer – the main focus is laid on the search of content, other functionalities like organizing & structuring content are neglected.

The keyframe based GUIs suggested by [6] are already implemented in a video database system for more than 2 years. This approach introduces some new concepts in the direction which keyframes should be displayed in connection with a summary (e.g. clustering of shots, giving importance scores for shots). Besides only displaying keyframes VIZARD introduces a new concept of efficiently summarizing videos, the so called O-M (object motion) images explained in chapter 3.2.

Corridini, Del Bimbo et al propose in [7] a new conceptual model for describing films together with a hypermedia navigation system to search and browse in the digital movies. The system is based on a feature extraction engine which offers shot detection, a simple scene analysis and a spatial temporal logic extraction. The resulting browsing interface is quite complicated and only meant for people who have deep background in image processing. In comparison to the proposed VExplorer the system is not dedicated to typical home users.

In [8] Hwang et al introduce video browsing features in connection with distance learning. The system offers students to make cross references, to link temporal-spatial sequences in videos with information (hyper-video links similar to [4] indicated via highlighted areas within the video), a synchronization facility allowing to view additional information synchronically with the video playback and database queries. The system is optimized for low bandwidth transmissions of the video (H.263 for video and G.723 for audio is used), hence only low quality video is stored and maintained within the system. Also because of the focus on education the system does not offer facilities like story sketching, which is offered by the proposed VExplorer.

The Informedia digital video library offers a number of useful features in order to search for video content [9]. One of those features is extracting geographic references from the video stream and link this information with maps in order to give the user the possibility to track the places related to the video content. Vice versa the user can also query for video material dealing with certain geographic areas. These geographic references are generated by utilizing a speech transcription engine applied on the audio track of the video in conjunction with natural language processing. The resulting metadata are stored together with the video essence and are used for queries. Despite this is a very interesting method, the Informedia system is very complex and thus out of scope for typical home users, which are the target audience of VIZARD. Informedia goes much beyond their typical needs. Also the system is – in contrast to the VExplorer – only suited in order to store and maintain video data and not in general multimedia data.

Core Media Asset Management systems (MAM) such as media archive® [15], Video Logger® [18], Ascential [16], Digital Library [17] are all large – in some cases even distributed – systems offering functionality focused on to manage tera and peta bytes of video data. Because these MAM target a completely different user group than the VExplorer and are therefore out of scope does we did not include a detailed comparison of these systems with the VExplorer in this paper.

3. THE VIZARD-EXPLORER

3.1 THE VIZARD FRAMEWORK

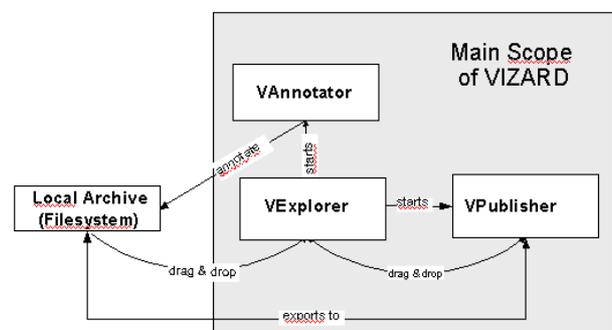The VIZARD Software will consist of three integrated modules given in Figure 1.

Figure 1: The VIZARD basic modules and their interaction

The VExplorer is used to import, organize, pre-process and browse multimedia material and sketch the storyline of a new video. The VAnnotator allows to generate annotations related to this material. Using the VPublisher enables to fine-tune the editing of the new movie. Both the VPublisher [13] and the VAnnotator [14] are developed by other partners of the VIZARD project and are not covered within this paper.

In this chapter we propose a tool – the VExplorer – which offers functionalities strongly related with the management and organization of multimedia content. The current appearance of the VExplorer is depicted in Figure 2. It shows the three main windows of the VExplorer: hierarchical tree-view (left window), list view (upper right window) and preview & navigation views (lower right window).
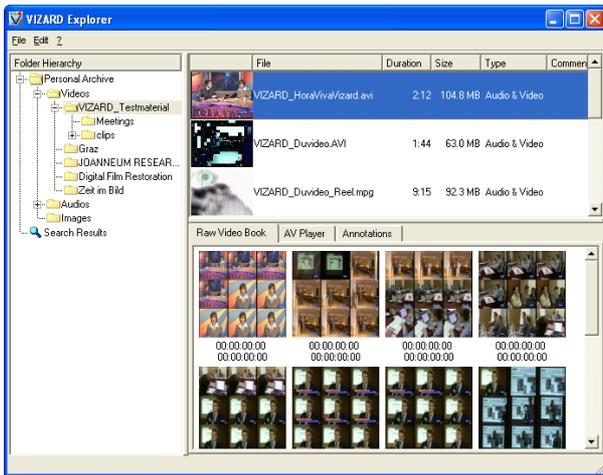
Figure 2: appearance of the VExplorer application

The VExplorer is designed to provide functionalities for managing any kind of MM data including video, audio, images and text. For managing multimedia data it will offer further functionalities such as outsourcing and a capture module. Besides the following functionalities will be offered:

- new navigation concepts within digital video essence (O-M images)

- interface for creating sketches

- search interface (content based retrieval & text)

- visual browsing tool for large archives

3.2 NAVIGATION AIDS

The VExplorer provides different navigation aids to quickly navigate through multimedia content:

- The *structure tree* allows to hierarchically structure the multimedia data in either a time-related manner or due to semantic context (e.g. combine all sequences about a certain topic into one group).

- The *list view* (right upper window of the VExplorer) displays either a single *keyframe* or in the thumbnail view three keyframes of the video in order to get a brief overview on the video content. Keyframes are extracted by a pre-processing module. These keyframes give a compressed overview about the content of the video.

- The *Object-Motion (O-M) images* give a very condensed summary of video essence [12]. O-M images are visualized in the preview & navigation view (lower right window of Figure 2). This data is further used by the VPublisher module for visualizing the video editing process.

- Two different types of *text annotations*: annotations covering the entire video and annotations covering parts of the video. For annotations covering the entire video text comments can be entered and later searched. For adding annotations covering only parts of the VAnnotator is started.

The O-M image is an even more compressed representation of the original video than keyframes are. It is composed of a motion image and two rows of object frames in two different sizes (see Figure 3) representing different temporal resolutions. The smaller object frames below one big frame represent the summary of content within the duration of time the big frame stands for.


Figure 3: O-M image with compressed keyframes

The motion image is created by adjoining the middle vertical column of pixels of every video frame (see Figure 4). This produces an image where the horizontal axis represents time dimension. This O-M image can be effectively used for getting very quickly information about video activity. Easily recognizable are e.g. shot boundaries (see Figure 3), sequences with low or little movement, zoom in and zoom out.
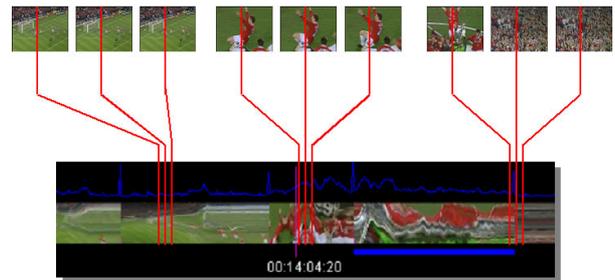
Figure 4: composition of a motion image

The VExplorer offers logical views on the physically content of the local hard disk or on network disks. As depicted in Figure 5 this allows to effectively structure multimedia content and prepare e.g. material to be used in a video editing software. In fact this feature is used in the VIZARD project exactly like this: the user pre-selects content for editing and uses then the VPublisher module – for editing the new video.
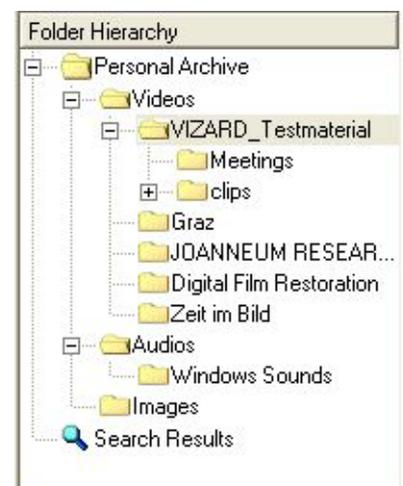
Figure 5: Hierarchy tree for logical views on content

3.3 SKETCHING A STORY

The VExplorer also offers an interface for creating sketches. This feature is actually based on the hierarchy tree window together with two other functionalities: rough cutting and rendering. The sketch creation interface is not fully developed yet, however, the functionality was already tested and implemented in an early software prototype called "VTreePlay" [11]. The basic idea is to support the creative process of creating a story with effective methods in the stages: selecting multimedia content suitable for the story, pre-structuring the story, selecting specific sequences from the pre-selected video essence and finally by rendering a preview video.
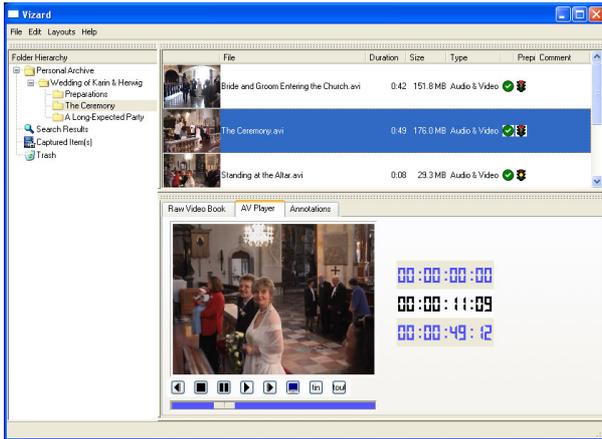

Figure 6: simple interface for creating sketches

Pre-selection of content can be done by using the hierarchy tree of the VExplorer. As an example we want to create a wedding video. In Figure 6, in a folder called "personal archive" all interesting content is first collected and structured by type.

The next step is to pre-structure the sketch. Following our example for the "Wedding video" we have created a separate folder "Herwig's Wedding" (see Figure 6) containing the scenes of the video we want to create. Note, that the folder structure now represents a temporal order, since the scenes are of course ordered temporally.

After pre-structuring the video essence containing the clips can be dragged & dropped from the "repository" to the corresponding scenes. Naturally only a short sequence is needed from the essence dropped into the folders. In order to cut out the needed clips a rough-cutting tool allows selecting a clip out of a longer sequence without needing to modify (re-encode) the essence itself (see Figure 7:).

Figure 7: Trimming of audio/visual media items

Trimming can done with the Buttons T_{in} and T_{out}. Clicking the buttons sets the time code of the selected segments to the current position of the video player. The time codes can also be set by setting the blue bar to the appropriate length with the mouse. The time code is shown right next to the player window (see Figure 6).

This trimming information will be stored in the VExplorer, however, the original essence will not be modified, hence this cutting is only virtual.

The final steps are to preview the created video and to generate a new video out of the selected clips. The preview can be simply performed by selecting the folder "wedding video" and clicking the play button under the player (see Figure 6) which automatically plays the content of the contained subfolders. Since all the folders represent a temporal order the final video will be played. When the user is satisfied with the video a new essence can be rendered by clicking the appropriate icon in the toolbar. All common formats are supported, e.g. MPEG-1/2, DivX, DV. For more sophisticated editing functionalities (transitions, effects, sound editing) the user would import the sketch into the VPublisher module.

3.4 SEARCH INTERFACE

The VExplorer supports two different kind of search methods:

- Text based search
- Visual, content based search

Text based search offers to search on previously generated annotations using the VAnnotator module. It is based on the video lens paradigm. A video lens represents one particular view on the material. Possible video lenses could describe technical description (media format, tape number, length) or content description (transcript, screenplay). The user is able to search within these annotations with a simple query dialog.

Content based search uses extracted features to visually compare a given image with video content. This comparison uses both image structures and color distribution. Color distribution analysis works well on "similar content" (e.g. footage from football games usually has a lot of green in it, a news anchor from any given TV station usually sits in front of the same background).

The matching items of a search are stored in a dedicated collection "search results" in the hierarchy tree (see Figure 5). Items from this collection can then be copied to any folder in the repository to be further used, e.g. in order to sketch a story (the user might have searched for content containing "brides" following our previous example).

3.5 VISUAL BROWSING TOOL

A problem of visualization arises when media collection get very large (500+ items in a folder including subfolders) An appropriate means of visualization is needed to get a quick overview.

Therefore a 2D/3D visualization of large multimedia collections is planned. Every item is represented by a single data point in a 2D or 3D space with the axes representing certain features (e.g. creation date, length). Features with a finite, discrete set of values (genre, director, actors) can be visualized using different shapes and colors of the data points

Especially in a 2D representation graphical queries can be made by selecting rectangular or circular areas of the data cloud.

3.6   IMPLEMENTATION DETAILS

The VExplorer uses the media handling functionality of Microsoft's DirectX/DirectShow libraries. Hence extension of media formats supported by the VExplorer can simply be achieved by installing a media codec available as DirectShow filter.

The multilingual GUI has been implemented using Trolltech's Qt library.

4.   OUTLOOK

The VIZARD project will be continuing until February 2003. For the rest of the project duration we plan to implement the following features into the VExplorer:

- Video capturing interface: A plug-in architecture open also for third party providers is planned. The GUI for controlling external devices shall be completely independent from the devices. As an example we want to develop a plug-in for capturing from DV devices.

- Repository export in MPEG-7 format: in order to make the VExplorer as open as possible to other systems we plan to develop an MPEG-7 export filter for the repository. This will allow to import the repository and the created stories into other applications.

- Link management: A challenging part of the VExplorer are the logical views realized via links to the physical essence. Deleting essence, however, may cause a lot of problems, e.g.: what to do with annotations related to the essence or the essence is already used in a VPublisher project. These are only two examples for many special cases which have to be solved.

- Outsourcing management: after some time every hard disk gets full especially when storing digital videos. For this case we plan to implement an outsourcing management supporting the user with saving and restoring data to a tape. At the same time all logical links will be kept and all data associated with it (annotations, usage in VPublisher projects, etc.) will be further maintained.

5.   CONCLUSIONS

We introduced the VExplorer, an efficient tool for visualizing, structuring, managing and browsing multimedia data. Capable of handling all kind of multimedia data the key features of the VExplorer are concentrated, however, on handling of audio/visual content. Innovative features of the VExplorer include:

- provision of efficient navigation aids such as hierarchical structure tree, O-M images, logical views and annotations.
- interface for creating sketches
- a search interface allowing for content based search as well as text based search

- a visual browsing tool for medium and large archives.

Further directions include implementation of a device independent capture module interface, repository export in MPEG-7, user friendly link management and an outsourcing management.

For the VExplorer an early prototype is already available as freeware and can be downloaded from [11].

6.   ACKNOWLEDGEMENT

REFERENCES

[1]   A. Munteanu, J. Cornelis, P. Cristea, Wavelet lossy and lossless image compression techniques – use of the lifting scheme, *Articles Digital Signal Processing Laboratory,* University "Politehnica" of Bucharest, Romania*, 2001, <URL: http://www.dsp.pub.ro/articles/wavletcomp/final.htm>*

[2]   S. Chang, W. Chen, H. J. Meng , H. Sundaram , D. Zhong, VideoQ, *Proceedings of the fifth ACM international conference on Multimedia* November 1997

[3]   D. L. Gall, MPEG, *Communications of the ACM* April 1991, Volume 34, Issue 4

[4]   H. Maurer, K. Andrews, F. Kappe, The Hyper-G Network Information System", J.UCS 1.4, pp. 206-220, 1995.

[5]   H. Maurer, J. Lennon, HyperWave: The Next Generation Web Solution", (Ed.), Addison-Wesley Longman, London, 1996.

[6]   A. Girgensohn, J. Boreczky, L. Wilcox, Keyframe-Based User Interfaces for Digital Video, *IEEE Computer Journal,* Vol. 34, No. 9; September 2001, pp. 61-67

[7]   J.M. Corridoni, A. Del Bimbo, D. Lucarella, Navigation and visualization of movies content, *Proceedings of the 11th International IEEE Symposium on Visual Languages*, 1995

[8]   H. Jenq-Neng, Y. Jeongnam, S. Deshpande, S. Ming-Ting, Video browsing for course-on-demand in distance learning, *Proceedings of the International Conference on Image Processing (ICIP '97) (3-Volume Set),* 1997

[9]   M. G. Christel, A. M. Olligschlaeger and C. Huang, Interactive Maps for a Digital Video Library, *Proceedings of the IEEE International Conference on Multimedia Computing and Systems Volume I,* 1998

[10] M. McQuaid, E. Booker, F. Chen, J. Holt, H. L.,
      Ana Lopez, Ch. Lyon, N. Shkarayeva, and L.
      Zhou, Vizard, Center for the Management of
      Information, University of Arizona, <URL:
      http://www.cmi.arizona.edu/research/kno_mgmt/
      vizard.htm>

[11] VTreePlay, Freeware developed for the VIZARD
      project by JOANNEUM RESEARCH, download
      from <URL: http://www.video-wizard.com>

[12] W. Haas, H. Müller, P. Uray, Visual Movie
      Annotation and Analysis in the VICAR Project,
      in: Jean-Yves Roger, Brian Stanford-Smith, Paul
      T. Kidd (Eds.), "Business and Work in the
      Information Society: New Technologies and
      Applications", pp. 382 - 387, IOS Press, 1999

[13] FH-JOANNEUM, Studiengang Informations-
      Design, <URL: http://informations-design.fh-
      joanneum.at/ >

[14] 4VDO, for video and multimedia, <URL:
      http://www.4vdo.pt/en/ie/4vdo.html>

[15] Blue order GmbH, media archive ®, <URL:
      http://www.media-transact.de/index.html >

[16] Ascential Software, Media360 ®, <URL:
      http://www.ascentialsoftware.com>

[17] IBM Corporation, Digital Library,
      <http://www.ibm.com>

[18] Virage Inc., VideoLogger ®,
      <http://www.virage.com>