

# New Interfaces for Video Browsing, Editing and Processing

Heimo Mueller, Dietmar Voelk, Konrad Baumann, Maja Pivec

Information Design Program

FH Technikum Joanneum University of Applied Science

A-8020 Graz, Austria

+43 316 5453 8600

{heimo.mueller, dietmar.voelk, konrad.baumann, maja.pivec}@fh-joanneum.at

## KEYWORDS

Video Editing, Video Browsing, Graphical User Interface, Visualization, Cognitive Models

## INTRODUCTION

This paper describes new user interface elements and interaction styles developed for video browsing, processing and editing.

In the first part of the paper we describe the use of OM images (OM = Objects and Movements) for the visualization of moving images. We will show how the technique of OM images can be the core element of a video browsing tool. This tool can be used for quick viewing and structuring of documentary video content (e.g. wedding videos, usability test videos, market research and educational videos).

Based on this we propose to redefine the video pre-production and video editing interface by applying the principles of typography to the description and manipulation of video objects. The so-called Video Book concept affords editing like text and therefore can be used intuitively by users familiar with today's word processing software.

We make use of OM images and the Video Book concept to create a graphical interface which leads users to a new mental model. This paper shows that when developing a new video publishing tool, we can learn from the long history of typography and book design.

## WHAT DO USERS WANT

In the field of video browsing, editing and processing, a lot of software tools are on the market. They are available in a wide range of price and features. Their user interfaces are based mostly on technology and follow the same concepts. Typical users are well trained professionals who have experience with these programs for years.

At FH Joanneum we have done some user studies who showed us that there are limitations, however, as these software products are about to enter the home market. Home users need self-explanatory products who do not

need much training. As video cameras have become popular during the last years, there is a need for easy-to-use video editing software for the home market. We generated a concept that offers a radically new approach for video editing user interfaces.

## WHAT IS A VIDEO BOOK

Most of us know how to write or edit text with a word processor, using line breaks, paragraphs, chapters and headings to structure the text content. Video content, however, is usually handled serially, which is more difficult to the users. So what would happen if we edit textual content according to the same principle as used in video editing? It would be serial information presented in a linear way as Figure 1 shows in a provocative way.



Figure 1: A linear book would look like this

On a page we have a lot of layout information in addition to the plain text, e.g. font type, character size, paragraph shape, white space, page border, title elements, and graphical objects as shown in Figure 2. The meaning of layout elements, e.g. "a heading marks the beginning of a new chapter" or "italic font style marks a quotation" is well known to all literate people. As well, people who use word

processors know how to use these elements, a knowledge which has spread across the western world during the last decades.

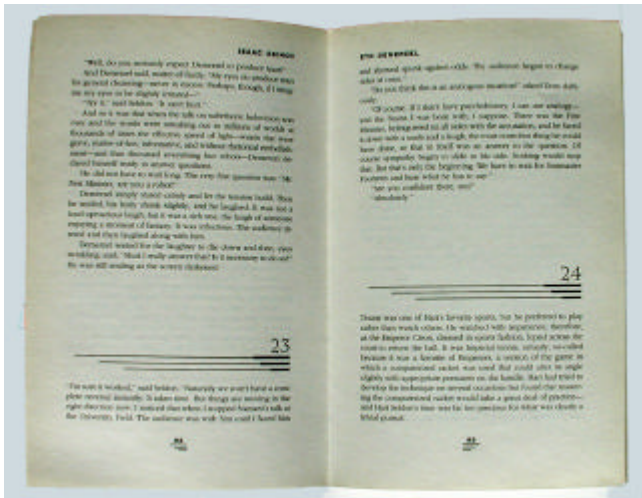


Figure 2: A sample book layout using paragraphs, white space, titles, page headings, page numbers and white space

From this comparison we conclude that books are represented in way that is richer and more easily graspable to the user than video content. Therefore we follow the idea of applying the principles of typography to the description and manipulation of video objects. In a first step shown in Figure 3 we break the linear video time-line into a number of lines (left) and add some lay-out information (right side).

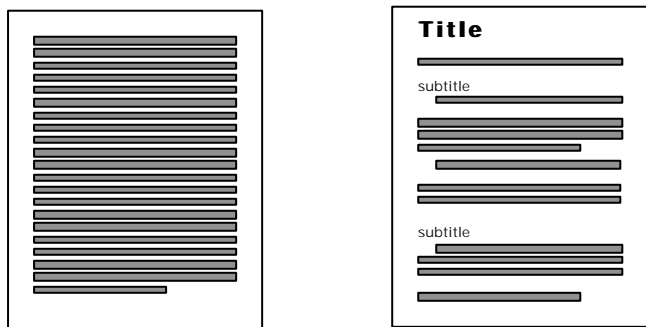


Figure 3: Breaking the time-line in video content (left) and adding meta-information (lay-out, right side) (Müller & Völk, 2000)

### OM - IMAGES

In order to break a timeline we have to find the points in time where to interrupt the continuous stream of images. A video consists of shots, i.e. sequences that have been recorded continually. A natural structure of video content is to break at shot borders and to hyphenate a shot at a fixed

temporal grid, e.g. every 4 seconds. The keyframes (pictures) that form the grid are called objects or O-images.

In addition we use a visualization of the movement called M-image, defined by the plane of maximum image flow in a 3-dimensional image volume. As Figure 4 shows, an M-image consists of one vertical line taken from every frame in a video and represents a time section across the video, orthogonal to the O-images or keyframes.

The units generated by this subdivision of the timeline are called content elements. They consist of a series of large O-images (larger time grid, time difference = 4 seconds), a series of small O-images (time difference = 1 second), and a corresponding M-image. All three representations of the video are superimposed. Browsing across this content element is easy and intuitive, as the user can find a specific position in the video either by looking at the keyframes (O-images) or at the continuous time scale visualized by the M-image. Figure 5 shows an example content element taken from the movie "The Wizard of Oz".

The technically interested reader can find further information in Müller & Tan (2000).

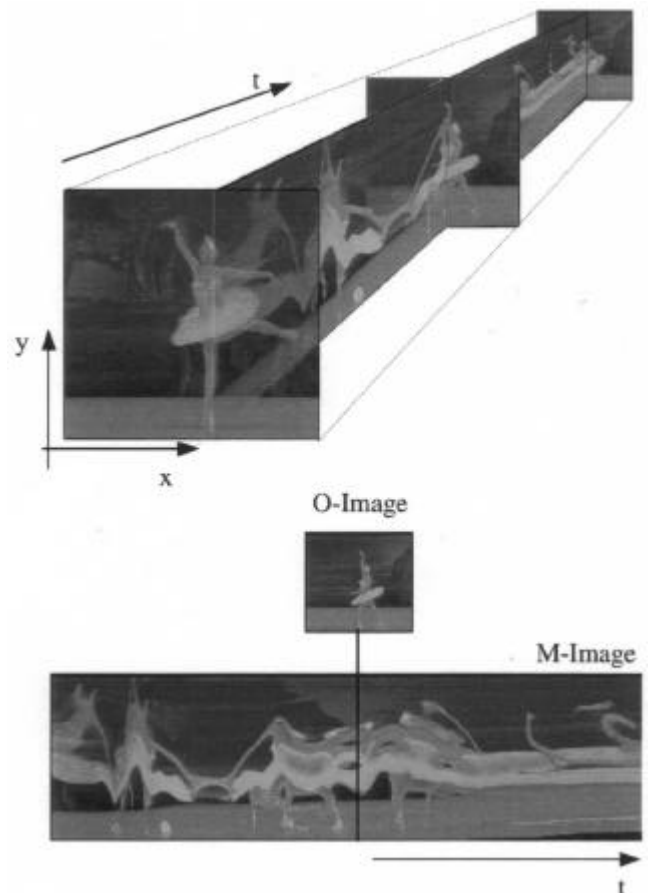


Figure 4: A visual representation of O-images (keyframes of the video) and an M-image (time section of the video); (Müller & Tan, 2000)



Figure 5: An example content element consisting of large O-images (top level), small O-images and an M-image (bottom level) (Müller & Völk, 2000)

A content element is the basic building block representing a piece of video which can be manipulated by the user. Between text editing and video editing we can identify the following equivalences:

- character = frame
- word = sequence of frames
- sentence = shot, clip
- paragraph = scene
- section = episode
- chapter = act

A content element is usually equivalent to a shot or clip.

If we use content elements in addition to text and graphics we come to the concept of a Video Book. This is a book formed of content elements and lay-out information, e.g. headings, subheadings and line breaks. Figure 6 shows an example "sentence" from the "Vizard of Oz" movie and a draft lay-out of a Video Book.



Figure 6: Video sentence consisting of video content elements (right side), Video Book (left side)

We can use the Video Book concept either for the visualization of the structure of an existing video and/or to support the whole creative process in video production. The definition of the first idea, a visual or textual sketch, the development of a script, the structuring of footage, editing, reviewing and documentation of the production process can be supported by the Video Book idea.

In a printed version of a Video Book, a director could even apply simple correction marks similar to the editorial marks used on a printed text document to communicate desired changes to the video production department. Figure 8 shows correction marks on a printed video sentence that consists of video content elements.



Figure 7: Correction marks in a Video Book (Müller & Völk 2001)

The Video Book concept allows home and professional users to start right at the beginning of the creative process when writing a video outline and script. The same tool can be used later when structuring of video content and video editing. For consumers the Video Book concept supports video document styles and predefined video elements (templates) in order to enable a simple and very intuitive creation of audio-visual content.

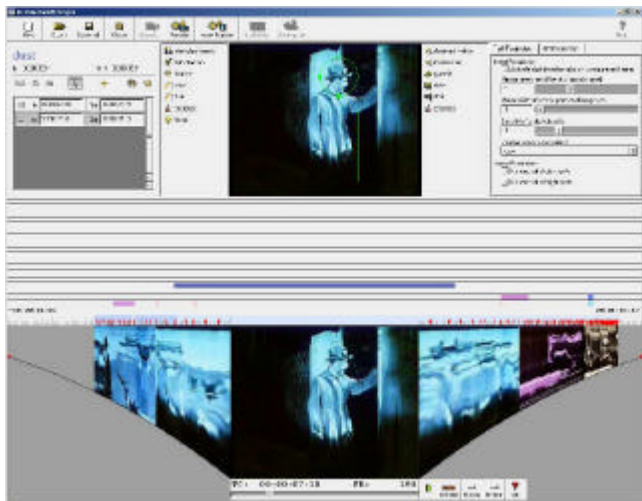
### USER INTERFACE CONCEPTS

Several user interface concepts have been created using OM-images. Figure 8 shows an interactive browser for movie and video content. In the lower middle part of the window there are three O-images and one M-image which is zoomed with a lens effect around the current position of the video cursor.



Figure 8: User interface concept for a video browser (Müller & Tan, 2000)

Another user interface developed at Technikum Joanneum is dedicated to the processing of old and damaged movie copies. It has an O-image in the center and M-images of the video content before and after the current position to both sides of the O-image (see Figure 9).



1998)

In the Vizard project a user interface based on the Video Book idea has been developed. The Video Book can be written and edited using pre-defined and new content elements. It is shown in Figure 10. The big window shows the Video Book looking like a page view of a text document. The small window is used to display video content. Tools and menus are used like in word processors.



Figure 10: User interface for Video Book editing (Müller &

## WHAT NEXT

First quick-and-dirty tests with the Video Book concept are promising, but still have to be done in-depth. The Video Book user interface is still late-breaking news and will undergo some refinement after the submission time of this paper. The ultimate goal which we will pursue in the next future is to meet user needs and to propose an entrance-level tool for video editing, helping to make this technology as simple and understandable as word processing.

## ACKNOWLEDGMENTS

The work reported here was carried out under the Training and Mobility of Researcher (TMR) program of the European Union. Our thanks are due to Marten den Uyl, and Herre Kuipers of Sentient Machine Research, Werner Haas and Peter Uray of Joanneum Research for their critical reviews and various discussions, and Tom Bauer of Technikum Joanneum. Furthermore we like to thank MIRAMAX for their support of the project.

## REFERENCES

- [1] Flickner J.P., Sawhney H., Niblack W., et al, Query by image and video content: the QBIC system, IEEE Computer, 28, September 1995 pp. 23-30.
- [2] Aigrain P., Zhang H.J., Petkovic D., Content-Based Representation and Retrieval of Visual Media – A State-of-the-Art Review, „Multimedia Tools and Applications 3(3), 1996, pp. 179-202.
- [3] Bolle R.M., Yeo B.-L., Yeung M.M. Video Query: Research directions, IBM Journal of Research and Development, Vol 42. No. 2 – Multimedia Systems.
- [4] Müller H., Haas W., Uray P., VICAR – An Automatic Video Annotation System, Proc. of EMMSEC 98, Bordeaux, Oct. 1998.
- [5] Minami K, Akutsu A, Hamada H., Video Handling with Music and Speech Detection, IEEE Multimedia July-September 1998, pp. 17-25.
- [6] Beigi M., Benitez A., and Chang S.-F., MetaSEEK: A Content-Based Meta Search Engine for Images, SPIE Conference on Storage and Retrieval for Image and Video Database, Feb. 1998.
- [7] Lienhart R., Kuhmünch C., Effelsberg. W. On the Detection and Recognition of Television Commercials, Proc. IEEE Conf. on Multimedia Computing and Systems, Ottawa, Canada, pp. 509-516, June 1997.
- [8] Zhang H.J., Tan C.Y., Smoliar S.W., Gong Y., Automatic Parsing and Indexing of News Video, Multimedia Systems, Vol. 2, No. 6, 1995, pp. 256-265.
- [9] Eakins J.P., Retrieval of trade mark images by shape feature, Proc of ELVIRA, 1994.

- [10] Herodotou N., Plataniotis K.N., Venetsanopoulos A.N., A Content-Based Storage and Retrieval Scheme for Image and Video Databases, *SPIE* Vol. 3309, pp. 697-708.
- [11] Tamura H., Mori S., Textural Features Corresponding to Visual Perception, *IEEE Trans. on Systems, Man and Cybernetics* Vol. 8, No. 6, 1978.
- [12] Akutsu A., Tonomura Y., Hashimoto H., Ohba Y., Video Indexing using Motion Vectors, *Proc. Visual Communications and Image Processing*, *SPIE*. Vol. 1818, 1992, pp. 1522-1530.
- [13] Ardizzone E., Cascia M.La, Video Indexing using Optical Flow Field, *Proc. of ICIP 96*, vol 3, 1996, pp. 831-834.
- [14] Kato T., Database Architecture for Content Based Image Retrieval, *Proc. of SPIE Conference on Image Storage and Retrieval Systems*, Vol. 1662, 1992, pp. 112-123.
- [15] Nam J., Tewfik A.H., Motion-based Video Object Indexing using Multiresolution Analysis, *SPIE* Vol. 3309, pp. 688-696.
- [16] Sahouria E., Zakhor A., Motion Indexing of Video, *Proc of ICIP 97*, vol. 2, 1997, pp. 526-529.
- [17] Uray P., Müller H., Plaschzug W., Haas W.: Visualising Artefacts, Meta Information and Quality Parameters of Image Sequences, *Proc. of the IS&T Conference on Visual Data Exploration and Analysis V*, San Jose, pp. 145-152, 1998.
- [18] Bordwell D., Thompson K., *Film art – an introduction*, 5<sup>th</sup> ed. McGraw-Hill, 1997.
- [19] Faulstich, W., *Einführung in die Filmanalyse*. 2te Auflage. Tübingen: Narr, 1980.
- [20] Zhong D., Zhang H., and Chang S.-F., Clustering Methods for Video Browsing and Annotation, *SPIE Conference on Storage and Retrieval for Image and Video Database*, Feb. 1996.
- [21] van Dijk T.A., *Macrostructures: An interdisciplinary study of global structures in discourse, interaction and cognition*. Hillsdale NJ, L. Erlbaum, 1980.
- [22] Arman et al., Content-Based Browsing of Video Sequences, *Proceedings of ACM Multimedia 94*.
- [23] Yeung M. et al., Video Browsing using Clustering and Scene Transitions on Compressed Sequences, *Multimedia Computing and Networking*, San Jose, Feb. 1995.
- [24] Müller, H., Tan, E. S., Visualizing the Semantic Structure of Film and Video, in *Visual Data Exploration and Analysis*, Robert F. Erbacher, Philip C. Chen, Jonathan C. Roberts, Craig M. Wittenbrink, Editors, *Proceedings of SPIE* Vol. 3960, pp. 393-400, 2000.
- [25] Müller, H., Völk, D., *The Vizard Bible*, internal document of the Vizard consortium, 2000.