AXES-RESEARCH — A User-Oriented Tool for Enhanced Multimodal Search and Retrieval in Audiovisual Libraries

Peggy van der Kreeft
Kay Macquarrie
New Media/Innovation Department
Deutsche Welle
Bonn, Germany
peggy.van-der-kreeft@dw.de
kay.macquarrie@dw.de

Max Kemman
Martijn Kleppe
Erasmus School of History, Culture and Communication
Erasmus University Rotterdam
Rotterdam, the Netherlands
kemman@eshcc.eur.nl
kleppe@eshcc.eur.nl

Kevin McGuinness
Insight Centre for Data Analytics
Dublin City University
Dublin, Ireland
kevin.mcguinness@eeng.dcu.ie

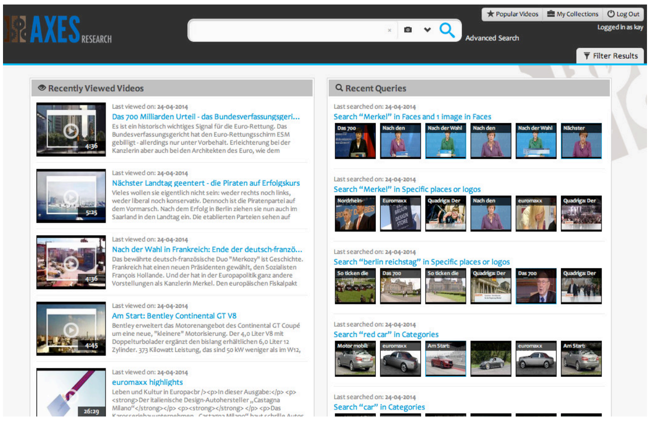
***Abstract*— AXES, Access for Audiovisual Archives, is a research project developing tools for new engaging ways to interact with audiovisual libraries, integrating advanced audio and video analysis technologies. The presented prototype is targeted at academic researchers and journalists. The tool allows them to search and retrieve video segments through metadata, audio analysis, as well as visual concepts and similarity searches. Presented here is a user-based vision on the research-oriented tool provided by AXES.**

***Keywords*— audiovisual archive; library; audio analysis; video analysis; similarity search; tool; information retrieval**

I. INTRODUCTION

The overall aim of AXES[1] (Access for Audiovisual Archives) and its consortium[2] is to develop tools that allow certain types of users to use digital audiovisual libraries in novel ways, helping them discover, browse, navigate, search and enrich archives. Three primary user groups are targeted: (1) archivists and broadcast professionals, dealing with content and metadata archiving and retrieval requests for reuse by media professionals; (2) the research community, including academic users and journalists looking for efficient tools for searching and retrieval of resource material; (3) the general home user in need of user-friendly ways to find interesting or entertaining audiovisual material in public repositories. In this paper we focus on the second user group, i.e. academic researchers and journalists.

II. AXES RESEARCH PROTOTYPE



Last viewed on: 24-04-2014
Das 700 Milliarden Urteil - das Bundesverfassungsgeri...
Es ist ein historisch wichtiges Signal für die Euro-Rettung. Das Bundesverfassungsgericht hat den Euro-Rettungsschirm ESM gebilligt - allerdings nur unter Vorbehalt. Erleichterung bei der Kanzlerin aber auch bei den Architekten des Euro, wie dem

Last viewed on: 24-04-2014
Nächster Landtag geentert - die Piraten auf Erfolgskurs
Vieles wollen sie eigentlich nicht sein: weder rechts noch links, weder liberal noch konservativ. Dennoch ist die Piratenpartei auf dem Vormarsch. Nach dem Erfolg in Berlin ziehen sie nun auch im Saarland in den Landtag ein. Die etablierten Parteien sehen auf

Last viewed on: 24-04-2014
Nach der Wahl in Frankreich: Ende der deutsch-franzö...
Das bewährte deutsch-französische Duo "Merkozy" ist Geschichte. Frankreich hat einen neuen Präsidenten gewählt, den Sozialisten François Hollande. Und der hat in der Europapolitik ganz andere Vorstellungen als Kanzlerin Merkel. Den europäischen Fiskalpakt

Last viewed on: 24-04-2014
Am Start: Bentley Continental GT V8
Bentley erweitert das Motorenangebot des Continental GT Coupé um eine neue, "kleinere" Motorisierung. Der 4,0 Liter V8 mit Doppelturbolader ergänzt den bislang erhältlichen 6,0 Liter 12 Zylinder. 373 Kilowatt Leistung, das sind 50 kW weniger als im W12,

Last viewed on: 24-04-2014
euromaxx highlights
Leben und Kultur in Europa<br /><p>In dieser Ausgabe:</p> <p><strong>Der italienische Design-Autohersteller „Castagna Milano“</strong></p> <p><strong></strong> </p> <p>Das Karosseriebauunternehmen „Castagna Milano“ baut schrille Autos

Last searched on: 24-04-2014
Search “Merkel” in Faces and 1 image in Faces

Last searched on: 24-04-2014
Search “Merkel” in Specific places or logos

Last searched on: 24-04-2014
Search “berlin reichstag” in Specific places or logos

Last searched on: 24-04-2014
Search “red car” in Categories

Last searched on: 24-04-2014
Search “car” in Categories

Fig.1 The AXES RESEARCH start interface provides users with a simple search interface and shows their recently viewed videos and queries.

[1] http://www.axes-project.eu

[2] ERCIM (France), KU Leuven (Belgium), Airbus Defence and Space (France), Technicolor (France), INRIA (France), Fraunhofer IAIS (Germany), Dublin City University (Ireland), University of Oxford (UK), University of Twente (the Netherlands), Netherlands Institute for Sound and Vision (the Netherlands), Erasmus University Rotterdam (the Netherlands), BBC (UK), Deutsche Welle (Germany)

978-1-4799-3990-9/14/$31.00 ©2014 IEEE

The AXES system is an advanced search and retrieval software that combines various technologies from computer vision such as face, object, and place recognition, similarity searching and automatic speech recognition, making it easier for the user to find relevant material, not dependent on available metadata. At this stage it is not a commercial product, but a prototype under construction.

So far two prototypes are available within the consortium. The first one, AXES PRO [1], was developed during the first and second years and was tested by media professionals. Using the same underlying system, and adding functionalities and offering a more user-oriented user interface, AXES RESEARCH was created and was tested by the research group at the end of 2013. This second prototype provides access through a user-friendly interface (see figure 1 above) specifically designed for the research community based on different prioritizations of user requirements [2]. The user interface offers the user different modes with varying levels of search, display detail, and user interaction.

III. TECHNOLOGIES COMBINED

The strength of the system is a combination of different technologies working in the background. The user interacts with the AXES system through a web-interface. The system consists of different modules, including a centralized link management and structured search system (for text indexing, maintaining links, fusing search results). Other modules (for audio searching, visual concept classification, similarity searches) use distributed servers and are linked with the central management system through web-service interfaces.

A wide range of advanced audio and video analysis technologies is used, including multimodal analysis of people, places, objects and events. The system can find people categories (e.g. a woman) or individuals (Angela Merkel); location categories (countryside) or specific places (Eifeltower); object categories (car) or specific objects (Mercedes logo or Italian flag); and event categories (mountain climbing). The analysis technologies are smoothly integrated with advanced linking technologies for deeper exploration of the content [6].

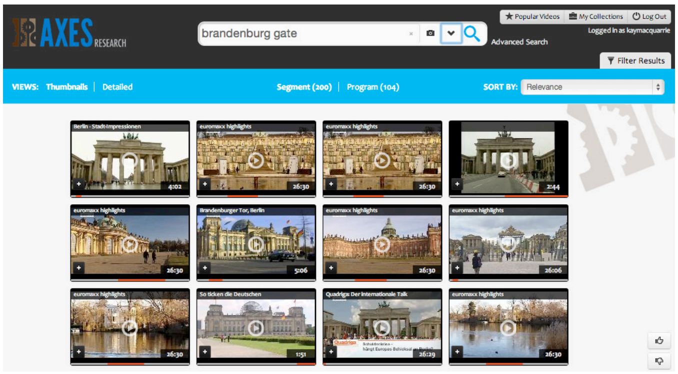
The following describes the key search technologies in more detail:

A. *Text Search / Spoken Words Search*

All metadata and spoken words are stored and indexed. Spoken words are provided in the form of a transcript originally provided by the content provider or they are automatically produced by Automatic Speech Recognition (ASR). The word error rate of ASR currently lies under 23% for German and English language and is expected to be even stronger reduced to under 18% for German audio with the use of neural networks and Gaussian mixture models. Users can search the textual data with standard text queries using Boolean conjunctives, such as AND and OR.

B. *Visual Search*

The system uses text-based queries to look for visual objects. This is done in conjunction with an external search engine and uses on-the-fly methods [3]. If a user makes a text search for "Brandenburg gate", the query is sent to a search engine like Google or Bing that produces a sample of the top-n images. From the results a model is created and used to detect similar objects in the archive. Queries might take some time (several seconds or longer) before a result is shown.


brandenburg gate

SORT BY: Relevance

Fig.2: AXES RESEARCH thumbnail view of visual search results. Results can also be viewed in detailed view.

To increase the responsiveness of the system, pre-trained objects were introduced, which include general and popular terms like house, cat or car. Additionally, new terms are learned and automatically added to the set of popular detectors, if a query (or a similar query) is posed repeatedly.

Generally, the system supports three types of visual search: visual categories [3], faces [10], and specific places or logos [4]. Further more, the user can also search for events. Events are "things that happen", e.g. a parade, protest march or an airplane taking off. The system recognises events based on multimodal input, including audio and visual features [5].

C. *Similarity Search*

Instead of entering keywords, a search can also be based on internal or external images. A similarity search can be done by using one or more images, either a keyframe shown from returned results, or an image uploaded by the user. Furthermore, the similarity search also supports user selectable search types. Currently, the user can select between faces and instances (an instance is basically anything but faces) to get the best results. Face similarity search uses a system based on facial landmarks. A set of 9 landmark points are detected, located on the eyes, nose, and mouth.

Instance search uses a specific search engine [9] to detect instances (i.e. the same scene or object, e.g. a house façade or book cover). The engine extracts interest points from a query using the Hessian Affine detector and computes a descriptor for each of them. It then searches in the database for images or scenes. The 300 most similar images are then matched with the query using a geometrical model. Images that pass this step are presented as results.

IV. AXES RESEARCH Features and Options

A. *Searching*

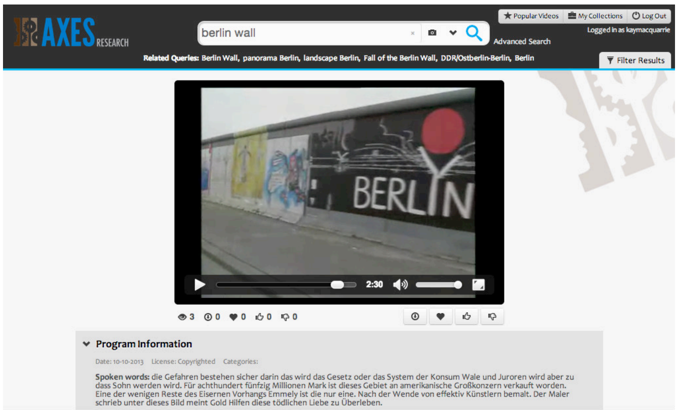
Users can choose to search in different ways. The first option is 'simple search', which is a single search bar where the user chooses whether to search on textual information, i.e. metadata or spoken words, or using a visual search, i.e. topical categories, faces, places or logos (see figure 2 above). Secondly, the advanced search mode allows for making combinations of several such queries, building up complex strategies. The third option is the "automatic search", which is most simple to the user as he or she only needs to provide keyword(s), leaving the system to select the most optimal search functionalities. The fourth option is a similarity search, for which users can upload an image or choose a thumbnail from a video within the AXES system to find similar videos. Another option for the user is to narrow down the results by filtering, using suggested filters (topic, name, place, publishing date, for example). A search on related videos and related segments is also possible.

B. *Deep links enable history of search trails*

The system provides deep links for most pages. This enables users to bookmark results, use the browser back and forward buttons, and share links with others. Moreover, deep links also memorize the type of searches that were used for a query. Thus, complex search trails can also later be accessed and adapted for a new search. Deep links are implemented for most of the major features, including queries, result lists, asset views, keyframe views, and preferences.

C. *Viewing, Saving, Downloading*

Depending on the amount of information the user wants to see in the results page, he or she has the option between different views, namely thumbnail (see figure 2 below) or detailed view where parts of associated metadata is shown. From the search results, the user can continue to asset view, i.e. full item display (see figure 3 below). Results can be downloaded, i.e. full video, videoclips, metadata, audio transcript, notes. The option "Personal collections" allows saving videos in separate personal collections, providing easy access to the search result strategy used and results for further selection and content gathering for future projects.



Related Queries: Berlin Wall, panorama Berlin, landscape Berlin, Fall of the Berlin Wall, DDR/Ostberlin-Berlin, Berlin

Program Information

Date: 10-10-2013 License: Copyrighted Categories:

Spoken words: die Gefahren bestehen sicher darin das wird das Gesetz oder das System der Konsum Wale und Juroren wird aber zu dass Sohn werden wird. Für achthundert fünfzig Millionen Mark ist dieses Gebiet an amerikanische Großkonzern verkauft worden. Eine der wenigen Reste des Eisernen Vorhangs Emmely ist die nur eine. Nach der Wende von effektiv Künstlern bemalt. Der Maler schrieb unter dieses Bild meint Gold Hilfen diese tödlichen Liebe zu Überleben.

Fig. 3 The asset view showing a single video or a segment and associated metadata. Audio transcripts, user notes and comments, and related videos are displayed further down the page.

D. *Editing*

On specific request of the research user group, a virtual cutter [7] is included as an editing tool, so that not only full videos, but also parts of the video can be selected and saved as customised videoclips. This makes the audiovisual material much more useful for subsequent compilation or detailed selection.

E. *Social Features*

Social user interaction is part of the tool, i.e. users can like, unlike or mark videos as favourite, and see such results from either themselves or other users: the most liked/unliked, favourited and most viewed videos. This can guide users for further searches.

Annotation is another feature which was specifically requested by researchers: users can read, add, and search private or public notes containing keywords, comments, reminders, etc. This is of particular interest to the academic community, who can use the tool for longer-term projects, for which efficient annotation of obtained results is essential for later reference and collaboration with or communication to team members or other interested parties [8].

V. User Testing

A total of 78 participants were involved in the evaluation sessions of AXES RESEARCH, including 48 journalists and scholars. In addition, the search engines and components have been demonstrated and tested at TRECVid and MediaEval benchmarking events during the course of the project [11]. Overall, participants were very interested in testing the new functionalities that could help them in their search endeavour. In general, the look and feel of the prototype was appreciated and users concluded that the functionalities of integrating video and audio, including similarity search, worked. User input and suggestions for enhancement serve to improve the coming versions of the AXES system and user interfaces.

VI. Outlook

AXES offers the user a novel way of exploring audiovisual content and interacting with this content in a personalised way. User interaction and user friendliness is ensured through the AXES user interfaces, with the two existing prototypes, i.e. AXES PRO and AXES RESEARCH, as well as the envisaged AXES HOME. The underlying system is the same for all AXES prototypes, but the user options and interfaces differ based on the established user requirements, resulting in an optimized user experience. End users can take advantage of a powerful system, without the need to be involved in all the technical intricacies. The different user interfaces, targeting different user groups, as well as the modular structure of the underlying system show potential for this enriched multimodal search and retrieval system.

Acknowledgment

This work is supported by the EU FP7 programme as EU Project FP7 AXES ICT-269980.

References

[1] K. McGuinness et al., “The AXES PRO video search system,” IEEE International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), July 2013.

[2] D. Nadeem, R. Ordelman, R. Aly, and E. Verbruggen, “Users Requirements in Audiovisual Search: A Quantitative Approach,” Research and Advanced Technology for Digital Libraries, 2013, pp. 241–246.

[3] O. M. Parkhi et al., “On-the-fly Specific Person Retrieval,” International Workshop on Image Analysis for Multimedia Interactive Services, 2012.

[4] B. Fernando, T. Tuytelaars, “Mining Multiple Queries for Image Retrieval: On-the-Fly Learning of an Object-Specific Mid-Level Representation,” ICCV 2013.

[5] Jérôme Revaud, Matthijs Douze, Cordelia Schmid, Hervé Jégou, “Event retrieval in large video collections with circulant temporal encoding,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2013, Portland, USA.

[6] Maria Eskevich et al., “Multimedia information seeking through search and hyperlinking,” Proceedings of the 3rd ACM International Conference on Multimedia Retrieval, Dallas, U.S. 2013.

[7] A. Rosendaal and J. Oomen, “The Davideon Project: Capitalizing the Possibilities of Streaming Video as Flexible Learning Objects for the Humanities,” Innov. J. Online Educ., vol. 2, no. 1, 2005.

[8] F. de Jong, R. Ordelman, and S. Scagliola, “Audio-Visual Collections and the User Needs of Scholars in the Humanities: a Case for Co-Development,” Proceedings of the 2nd Conference on Supporting Digital Humanities (SDH 2011), 2011.

[9] H. J´egou, M. Douze, and C. Schmid, “Improving bag-of-features for large scale image search,” International Journal of Computer Vision, vol. 87, no. 3, pp. 316–336, 2010.

[10] K. Simonyan, O. M. Parkhi, A. Vedaldi, A. Zisserman, “Fisher Vector Faces in the Wild” British Machine Vision Conference, 2013

[11] M. Eskevich, R. Aly, R. Ordelman, S. Chen and G. J. Jones, “The Search and Hyperlinking Task at MediaEval 2013,” Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop, 2013.