**KLEPPE — AXES**

<2 figures>

**AXES: Researching and Accessing Videos through Multimodal Analyses**

**Kleppe, M., Kemman, M., Van der Kreeft, P., Macquarrie, K. and McGuinness, K.**

AXES—Access for Audiovisual Archives is a research project developing tools for new and engaging ways to interact with audiovisual libraries, integrating advanced audio and video analysis technologies. This poster and demo will showcase the system that is developed for academic researchers and journalists. The tool allows them to search and retrieve video segments through metadata and audio analysis, as well as visual concepts and similarity searches.

**Background**

In the near future, audiovisual material is perhaps the biggest wave of data that will become available for academic researchers (Smith, 2013). This type of material has a big value for the digital humanities as it is multilayered: a single document can provide information regarding language, emotions, speech acts, narrative plots, and references to people, places, and events. This richness provides interesting data for various disciplines and holds the promise of multidisciplinary collaboration between, e.g., computer sciences, social sciences, and the humanities (Ordelman et al., 2014).

However, the use of audiovisual data by scholars in the humanities and the application of digital methods for analysis are still in their infancy for several reasons. One of them is the lack of useful systems that help academic researchers search through the large amounts of audiovisual data (De Jong et al., 2011). Given the multimodal nature of audiovisual data, different types of techniques are required to provide access to the data. The overall aim of the research project AXES—Access for Audiovisual Archives is to develop tools that allow novel ways of using digital audiovisual libraries, helping users to discover, browse, navigate, search, and enrich archives.1 Within the project, three systems are developed. In this poster and demonstration we will show the AXES RESEARCH system that was developed to cater the needs of humanities scholars and journalists.

**AXES RESEARCH System**

Building on several requirements studies amongst humanities scholars (Kemman et al., 2014; 2013a; 2012) and journalists (Kemman et al., 2013b), the AXES RESEARCH system is an advanced search and retrieval system 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 (Van der Kreeft et al., 2014). The strength of the system is a combination of different technologies working in the background.



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

The following key search technologies are used in the AXES RESEARCH system: text/spoken words search, visual search, and similarity search.

*Text Search / Spoken Words Search*

All metadata of the audiovisual programs and its spoken words are stored and indexed. Spoken words are provided in the form of a transcript originally provided with the audiovisual data or are automatically produced by Automatic Speech Recognition (ASR).

*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 (Parkhi et al., 2012). 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.



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

Generally, the system supports three types of visual search: visual categories (Parkhi et al., 2012), faces (Simonyan et al., 2010), and specific places or logos (Fernando et al., 2013). Furthermore, the user can also search for events. The system recognizes events based on multimodal input, including audio and visual features (Revaud et al., 2013).

*Similarity Search*

Instead of entering keywords, a search can also be based on internal or external images, also known as content-based image search (Smeulders, 2000). 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, comparable with the query by image technique as implemented at Google Images.2

**Results: User Testing**

A total of 78 participants were involved in the evaluation sessions of AXES RESEARCH. Overall, participants were very interested in the system that assists them in research practices. 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 served to improve the coming versions of the AXES system that will focus on home users.

**Conclusion**

AXES RESEARCH offers academics a novel way of exploring audiovisual content. They can take advantage of a powerful system, without the need to be involved in all the technical intricacies, allowing them to incorporate audiovisual materials in their research practice, which is currently rarely done given the absence of a system like AXES RESEARCH that helps them search through large amount of audiovisual data.

**Acknowledgment**

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

**Notes**

1. www.axes-project.eu.

2. www.images.google.com.

**References**

**De Jong, F., Ordelman, R. and Scagliola, S.** (2011). Audio-Visual Collections and the User Needs of Scholars in the Humanities: A Case for Co-Development. In *Proceedings of the 2nd Conference on Supporting Digital Humanities (SDH 2011)*,Copenhagen, Denmark, 17–18 November 2011.

**Fernando, B. and Tuytelaars, T.** (2013). Mining Multiple Queries for Image Retrieval: On-the-Fly Learning of an Object-Specific Mid-Level Representation. *ICCV*, Sydney, Australia, 3–6 December 2013.

**Kemman, M., Kleppe, M., Beunders, H.** (2012). Who Are the Users of a Video Search System? Classifying a Heterogeneous Group with a Profile Matrix. In *13th International Workshop on Image Analysis for Multimedia Interactive Services*, Dublin, Ireland, 23–25 May 2012.

**Kemman, M., Kleppe, M. and Maarseveen, J.** (2013a). Eye Tracking the Use of a Collapsible Facets Panel in a Search Interface. In *Research and Advanced Technology for Digital Libraries*. Berlin: Springer, pp. 405–8, doi:10.1007/978-3-642-40501-3\_47.

**Kemman, M., Kleppe, M., Nieman, B. and Beunders, H.** (2013b). Dutch Journalism in the Digital Age [El Periodismo Holandés En La Era Digital]. *Icono 14,* **11**(2): 163–81, doi:10.7195/ri14.v11i2.596.

**Kemman, M., Kleppe, M. and Scagliola, S.** (2014). Just Google It. In Mills, C., Pidd, M. and Ward, E. (eds), *Proceedings of the Digital Humanities Congress 2012*. Sheffield, UK: HRI Online Publications.

**Ordelman, R., Kemman, M., Kleppe, M. and De Jong, F.** (2014). Sound and (Moving Images) in Focus: How to Integrate Audiovisual Material in Digital Humanities Research. *Digital Humanities 2014*, Lausanne, Switzerland, 6–12 July 2014, http://dharchive.org/paper/DH2014/Workshops-914.xml.

**Parkhi, O. M. et al.** (2012). “On-the-Fly Specific Person Retrieval.” In *13th International Workshop on Image Analysis for Multimedia Interactive Services*, Dublin, Ireland, 23–25 May 2012.

**Revaud, J., Douze, M., Schmid, C. and Jégou, H.** (2013). Event Retrieval in Large Video Collections with Circulant Temporal Encoding. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013, Portland, OR.

**Simonyan, K., Parkhi, O. M., Vedaldi, A. and Zisserman, A.** (2013). Fisher Vector Faces in the Wild. British Machine Vision Conference, Bristol, UK, 9–13 September 2013.

**Smeulders, A. W., Worring, M., Santini, S., Gupta, A. and Jain, R.** (2000). Content-Based Image Retrieval at the End of the Early Years: Pattern Analysis and Machine Intelligence. *IEEE Transactions,* **22**(12): 1349–80.

**Smith, J. R.** (2013). Riding the Multimedia Big Data Wave. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval—SIGIR ’13*. New York: ACM Press, doi:10.1145/2484028.2494492.

**Van der Kreeft, P., Macquarrie, K., Kemman, M., Kleppe, M. and McGuinness, K.** (2014). AXES-RESEARCH: A User-Oriented Tool for Enhanced Multimodal Search and Retrieval in Audiovisual Libraries. In *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*,IEEE, Klagenfurt, Austria, 18–20 June 2014, 1–4, doi:10.1109/CBMI.2014.6849852.