

Crowdsourcing Rock N' Roll Multimedia Retrieval

Cees G.M. Snoek
ISLA, Informatics Institute
University of Amsterdam
Science Park 107, 1098 XG
Amsterdam, The Netherlands

Bauke Freiburg
Video Dock
Panamalaan 1b, 1019 AS
Amsterdam, The Netherlands

Johan Oomen
Netherlands Institute for
Sound and Vision
Sumatrallaan 45, 1217 GP
Hilversum, The Netherlands

Roeland Ordelman
Human Media Interaction
University of Twente
Box 217, 7500 AE
Enschede, The Netherlands

ABSTRACT

In this technical demonstration, we showcase a multimedia search engine that facilitates semantic access to archival rock n' roll concert video. The key novelty is the crowdsourcing mechanism, which relies on online users to improve, extend, and share, automatically detected results in video fragments using an advanced timeline-based video player. The user-feedback serves as valuable input to further improve automated multimedia retrieval results, such as automatically detected concepts and automatically transcribed interviews. The search engine has been operational online to harvest valuable feedback from rock n' roll enthusiasts.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval: Information Search and Retrieval
General Terms: Algorithms, Experimentation, Performance
Keywords: Semantic indexing, video retrieval, information visualization

1. INTRODUCTION

Despite years of vibrant research, multimedia retrieval is often criticized for lack of real-world applications [5, 6]. Searching for video on the web, for example, is still based on (user provided) text. Automated alternatives using speech recognition are only sparingly used online, Blinkx being a notable exception. Less mature technology, like visual concept detection, is non-existent in real-world applications. To prevent complete dependence on automated analysis, and the associated errors, some suggest to exploit user-provided feedback for effective multimedia retrieval [9].

In this paper, we demonstrate a real-world video search engine based on advanced multimedia retrieval technology, which allows for user-provided feedback to improve and ex-

tend automated content analysis results, and share video fragments. To encourage feedback, we focus on a dedicated user community of rock n' roll enthusiasts from a Dutch rock festival. Our online search engine allows to retrieve fragments from forty years of rock n' roll video footage recorded during the festival, which was previously only available in an offline archive of digital cultural heritage. Different from existing work on concert video retrieval, which emphasizes visual concept detection [10], browsing [3, 8], or user-generated content organization [7], we consider the integrated usage of multimedia retrieval technology coupled with a simple, intuitive, and easy to use crowdsourcing mechanism the main contribution of this work.

2. CROWDSOURCING MECHANISM

2.1 Encouraging User Feedback

In order to find a balance between an appealing user experience and a maximized user participation, we motivate



Figure 1: Harvesting user feedback for video fragments. The thumbs-up button indicates agreement with the automatically detected label, thumbs-down indicates disagreement. Three keyframes represent the visual summary of the fragment.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'10, October 25–29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-60558-933-6/10/10 ...\$10.00.



Figure 2: Timeline-based video player with visual detection results marked by colored dots. In addition to standard video playback functionality, users can navigate directly to fragments of interest by interaction with the colored dots, which pop-up a feedback overlay as displayed in Figure 1.

online users to participate by providing them with access to a selection of exclusive, full-length concert videos. The users watch the videos without interruption and are encouraged to provide their feedback by graphical overlays that appear on top of the video, see Figure 1.

The threshold to participate is deliberately kept low. Users do not need to sign up and can provide their feedback just by clicking buttons. With the thumbs-up button they indicate that they agree with the automatically detected label for the video fragment. If they press the thumbs-down button, the user is asked to correct the label. Within a few clicks the user can select another pre-defined label or create a new label on demand. In addition, we allow the users to indicate whether the start or end of the fragment is inconsistent with the label. All user feedback is stored in a database together with the users IP addresses and user sessions.

2.2 Timeline Video Player

The main mode of user interaction with our video search engine is by means of a timeline-based video player, see Figure 2. The player enables users to watch and navigate through a single video concert. Little colored dots on the timeline mark the location of an interesting fragment corresponding to an automatically derived label. To inspect the label and the duration of the fragment, users simply move their mouse cursor over the colored dot. By clicking the dot, the player instantly starts the specific moment in the video. If needed, the user can manually select more concept labels in the panel on the left of the video player. If the timeline becomes too crowded as a result of multiple labels, the user may decide to zoom in on the timeline. Besides providing feedback on the automatically detected labels, we also allow our users to comment on the individual fragments, share the fragment through e-mail or Twitter, and embed the integrated video player, including the crowdsourcing mechanism, on different websites.

3. MULTIMEDIA SEARCH ENGINE

3.1 Rock N’Roll Video Archive

Our search engine uses archived video footage of the Pinkpop festival. This annual rock festival is held since 1970 at Landgraaf, the Netherlands. All music videos have been recorded during the 40 years life cycle of the festival. We cleared copyright for several Dutch and Belgian artists playing at Pinkpop, including gigs from *dEUS*, *Golden Earring*, and *Urban Dance Squad*. The amount of footage for each festival year varies from only a summary to almost unabridged concert recordings, even including raw, unpublished footage as well as several interviews with the artists. The online rock n’ roll video archive contains 32 hours in total.

3.2 Interview Speech Recognition

Automatic speech recognition (ASR) technology was used to attach browsing functionality to the interview fragments in the collection. Speech transcripts were generated using the SHoUT ASR toolkit [4] and post-processed to generate a filtered term frequency list that is most likely to represent the contents of the interviews, based on tf.idf statistics. This list was then used to create a time-synchronized term cloud. Each word in the cloud is clickable to enable users to jump to the part of the interview where a word is mentioned.

3.3 Visual Concert Concepts

In contrast to domains like news video, where the number of visual concepts is unrestricted, the number of concepts that may appear in a concert is more or less fixed. A band plays on stage for an audience. Thus, major concepts are related to the role of the band members, e.g., lead singer, or guitarist, and the type of instruments that they play, such as drums or keyboard. Although quite many instruments exist, most bands typically use guitars, drums, and keyboards. We

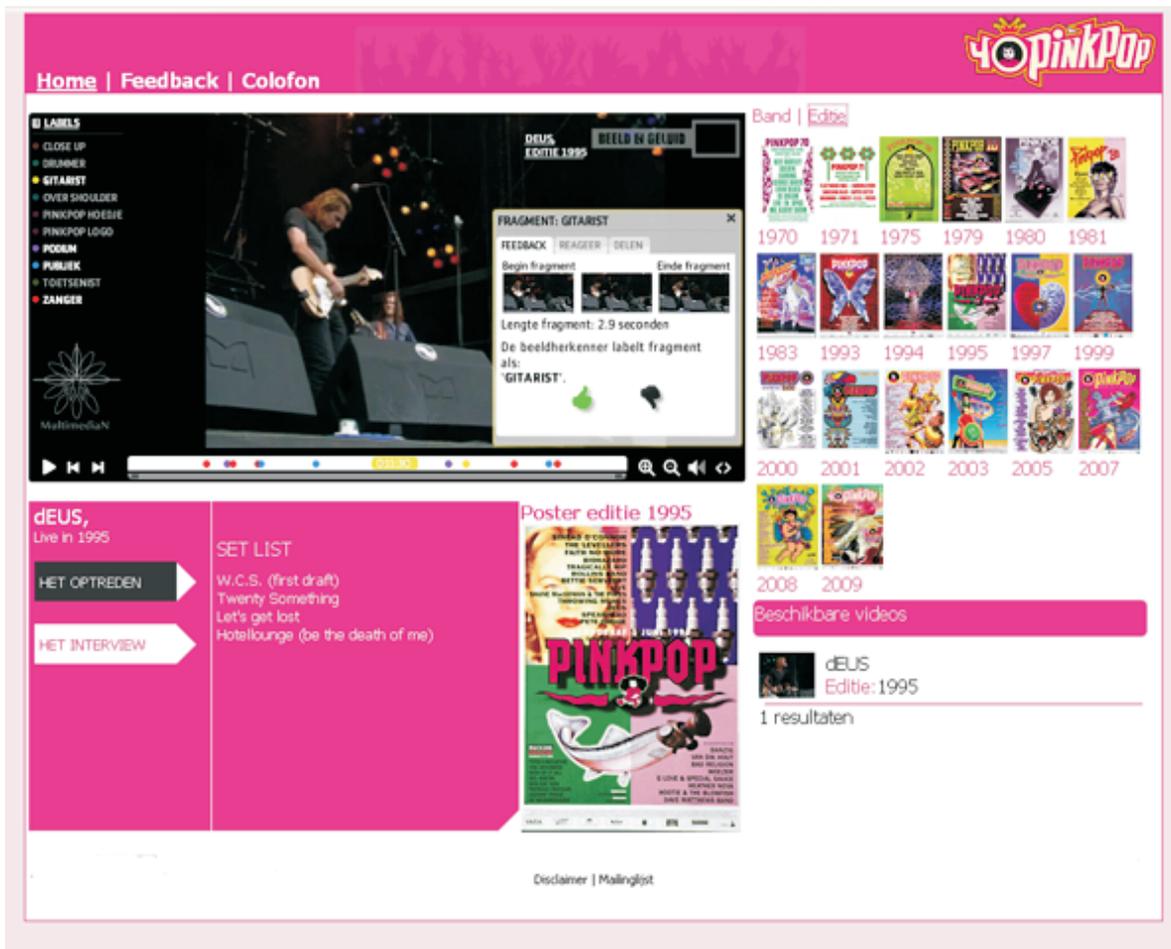


Figure 5: User interface of the multimedia search engine for rock n' roll video. Core component is the crowdsourcing mechanism (Figure 1), which relies on online users to improve, extend, and share, automatically detected results in video fragments using an advanced timeline-based video player (Figure 2). Users may select a concert using the navigation panel on the right. For retrieved concerts, meta-data such as the set list and an interview with the artists is highlighted. The search engine has been operational online to harvest valuable feedback from rock n' roll enthusiasts at <http://www.hollandsglorieoppinkpop.nl/> (includes video).

- [4] M. Huijbregts, R. Ordeman, L. van der Werff, and F. M. G. de Jong. SHoUT, the University of Twente submission to the n-best 2008 speech recognition evaluation for Dutch. In *Proceedings of Interspeech*, pages 78–90, Brighton, UK, 2009.
- [5] A. Jaimes, M. Christel, S. Gilles, R. Sarukkai, and W.-Y. Ma. Multimedia information retrieval: What is it, and why isn't anyone using it? In *Proceedings of the ACM SIGMM International Workshop on Multimedia Information Retrieval*, pages 3–8, Hilton, Singapore, 2005.
- [6] M. Kankanhalli and Y. Rui. Application potential of multimedia information retrieval. *Proceedings of the IEEE*, 96(4):712–720, 2008.
- [7] L. S. Kennedy and M. Naaman. Less talk, more rock: automated organization of community-contributed collections of concert videos. In *Proceedings of the International Conference on World Wide Web*, pages 311–320, Madrid, Spain, 2009.
- [8] S. U. Naci and A. Hanjalic. Intelligent browsing of concert videos. In *Proceedings of the ACM International Conference on Multimedia*, pages 150–151, Augsburg, Germany, 2007.
- [9] D. A. Shamma, R. Shaw, P. L. Shafton, and Y. Liu. Watch what I watch: using community activity to understand content. In *Proceedings of the ACM SIGMM International Workshop on Multimedia Information Retrieval*, pages 275–284, Augsburg, Germany, 2007.
- [10] C. G. M. Snoek, M. Worrning, A. W. M. Smeulders, and B. Freiburg. The role of visual content and style for concert video indexing. In *Proceedings of the IEEE International Conference on Multimedia & Expo*, pages 252–255, Beijing, China, 2007.
- [11] J. C. van Gemert, C. G. M. Snoek, C. J. Veenman, A. W. M. Smeulders, and J.-M. Geusebroek. Comparing compact codebooks for visual categorization. *Computer Vision and Image Understanding*, 114(4):450–462, 2010.