Int J Digit Libr (2004) 4: 42–44 / Digital Object Identifier (DOI) 10.1007/s00799-003-0065-5

Experiences of creating four video library collections with the Físchlár system *

Alan F. Smeaton, Hyowon Lee, Kieran McDonald

Centre for Digital Video Processing, Dublin City University, Glasnevin, Dublin 9, Ireland
e-mail: Alan.Smeaton@dcu.ie

Published online: 23 July 2004 –

Abstract. This paper describes how the FÍSCHLÁR system, which supports indexing, browsing and searching through archives of digital video information, has been used to create four separate video libraries of information. We briefly introduce FÍSCHLÁR and then describe its application in FÍSCHLÁR-TV (a digital library of recorded broadcast TV content, updated regularly), FÍSCHLÁR-News (a digital library of TV news, updated daily), and FÍSCHLÁR-Nursing (a digital library of video teaching materials in the domain of nursing), and how FÍSCHLÁR has also been used to provide searching through a collection as part of the TREC2002 video track interactive user experiments. Our experiences show that the range of user requirements for accessing video content seems to be much broader than for any other media, which makes the development of video access techniques very challenging.

Keywords: Hypertext – Path maintenance – Digital library services – Directed paths – Guided tours

1 The Físchlár digital video library system

Digital video information is becoming hugely commonplace, with content being generated from TV, movies, DVDs and CCTV, as well as by individual consumers with digital camcorders. While many developments and progress in video coding standards have helped to accelerate the production of digital video information, progress in the development of techniques to actually manage this content effectively has been much slower. In this short paper we report on our experience in developing applications for the FÍSCHLÁR system, which is used to manage video content in a variety of application scenarios.

* The support of the Informatics Research Initiative of Enterprise Ireland is gratefully acknowledged.

FÍSCHLÁR is an end-to-end digital library system which supports capture, indexing, storing, browsing, searching and summarising of digital video information in a range of applications. All applications share the same underlying architecture, which is designed to support XML-based processing of information. FÍSCHLÁR is accessed through a Web browser, and the interface internally produces responses for the user as XML documents which use XSL stylesheets to render device-dependent HTML for viewing on different platforms. We use a standard format for the representation and exchange of derived metadata, MPEG-7, which is an ISO standard for multimedia content description.

A fundamental component of FÍSCHLÁR is the analysis that the system automatically performs on incoming video information, normally in MPEG-1 format. In "traditional" video library environments as used in most TV archive libraries, content is hand-annotated at the shot level via text descriptions and these are then used as the basis for subsequent retrieval. The FÍSCHLÁR approach includes the automatic detection of shot boundaries and the selection of representative keyframes from shots above a certain length, techniques which are based on our extensive prior research and evaluation. In FÍSCHLÁR, a spoken transcript can be acquired either through the direct capture of teletext (closed captions) in the case of broadcast TV or through the use of speech recognition software in the case of other collections. FÍSCHLÁR also supports indexing of video material based on features automatically extracted from the video source by external processes, features such as discrimination between speech and music [3] and the recognition of on-screen faces [1]. Identified features such as these are marked up in MPEG-7 before incorporation into FÍSCHLÁR.

User information seeking with the FÍSCHLÁR system is based on finding and then selecting a video program through text or other metadata searching. Once a program has been selected, FÍSCHLÁR supports several in-

terfaces which allow rapid browsing through automatically extracted keyframes or keyframe summaries of that video. These include a keyframe slideshow, a hierarchical keyframe browser, and a timeline browser, as described in [4]. Once a section of one of the video programs in the library being searched has been identified as being of interest to the user, the user clicks on the appropriate keyframe and the video is streamed from that point forward in the video.

2 Físchlár video library collections

Físchlár has been used to support video navigation through four separate video collections, and each is described briefly.

2.1 Físchlár-*TV*

Físchlár-TV allows users to record, browse, search and watch TV programs broadcast from any of eight terrestrial TV stations in Ireland. The purpose is to provide a large-scale, communal TV recording and playback resource for both educational and entertainment purposes. Programs to be recorded are selected from an online TV schedule which can be personalised based on past preferences in recording and playback. Once recorded, programs are analysed to determine shot bounds and keyframes, indexed and placed in a communal video library for subsequent searching, keyframe browsing and playback (Fig. 1). At any point in time there are hundreds of recorded programs available, and each is automatically assigned one or more genres or classifications such as "movie", "soap", "sports" or "kids". These classifications are helpful for users in browsing the library contents in order to find programs.

Users of Físchlár-TV generally require access to entire TV programs to view as a whole rather than more fine-grained access to smaller units such as shots or scenes. While we have developed a text-based search of the closed captions of programs, this is not really used much to help users of Físchlár-TV find the content they are seeking.

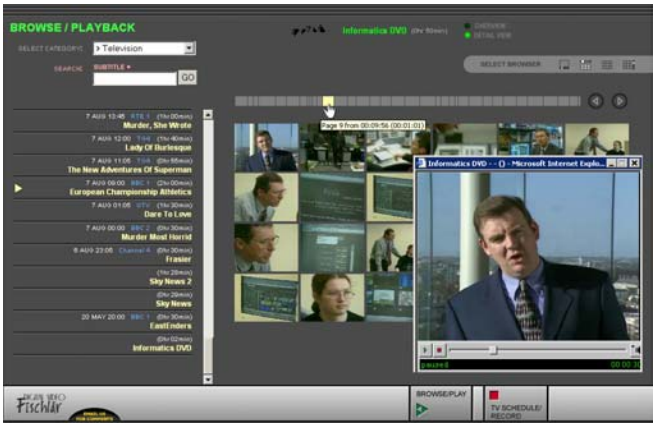

Fig. 1. Físchlár-TV browsing interface

2.2 Físchlár-*News*

The Físchlár-News system also records materials from broadcast TV, but in this case it is the main evening TV news from the national broadcast station in Ireland, RTÉ. Físchlár-News has built up an archive of hundreds of hours of news materials. The system is designed to be used as an aid to faculty and students of journalism and media on our campus to help them locate news stories and news clips about current affairs. At present, video material in Físchlár-News is indexed by a combination of keyframes derived from shots and text through captured teletext, and text searching is as important for navigation as keyframe browsing. Our experiences with Físchlár-News indicate that users want to access this content at the level of news *stories* rather than news *programs*, where there are usually between 10 and 20 news stories in each news program. Users like to use text searching or program selection by date to find a news story, play that story and then skip to the next story of interest to them, either a linked or related story, or a story on a completely different topic. If the 30-min daily broadcast news programs are segmented into individual news stories, these individual stories can be used as the unit of users' searching and browsing; thus, for example, the search result can be presented as a list of matched stories against a user's query.

Because of the specific information needs of users of Físchlár-News, we are extending this system to do automatic news story segmentation based on an analysis of the audio track to determine speech, speaker discrimination and music, and an analysis of the visual component to determine anchorperson, outside broadcast and advertisement detection [5]. When combined, these will allow us to break up a news program into constituent news stories and allow direct searching and retrieval of segmented news stories.

2.3 Físchlár-*TREC2002*

Físchlár-TREC2002 was developed to support our participation in the TREC2002 video track [2]. In this system, 35 h of video material have been indexed and are searchable via an automatic transcript of the speech as well as by the presence of automatically identified features such as faces or people, indoor or outdoor, landscape or cityscape, monologue, dialogue or music, and on-screen text, all represented in MPEG-7. This set of features has been chosen as representative of the kind of features which users wish to incorporate into their information needs as discriminators, but the feature set is representative rather than comprehensive. Once a user has run a query representing their information needs on

the set of descriptors, highly ranked shots are presented and grouped into video programs, and each shot is presented as a keyframe plus some iconic indicators of the presence of the features mentioned above.

Físchlár-TREC2002 is different from the other systems we have developed in that the unit of retrieval for users is the shot, almost the smallest unit of video that can be identified. The scenario which Físchlár-TREC2002 addresses is a user searching through a video library to identify a shot to illustrate some point in a documentary or news program, and as such this represents a real user information need. Once a relevant shot has been identified and logged, the user will want to find subsequent relevant shots, not by browsing through neighbouring keyframes as in Físchlár-TV or Físchlár-News, but by browsing through shots which are similar to the already-identified shot in terms of overlap of identified and unidentified features. Thus Físchlár-TREC2002 should ideally also support a "find more shots like this shot" search feature.

2.4 Físchlár-*Nursing*

The final Físchlár system we have developed is Físchlár-Nursing, which supports navigation through a collection of educational video materials for students of our School of Nursing, which forms part of their core taught curriculum. While users of the other Físchlár systems require fine-grained browsing and, in the case of Físchlár-News and Físchlár-TREC2002, extensive support for searching, users of Físchlár-Nursing require a combination of an initial viewing of the whole material followed by subsequent browsing and searching for specific clips at times of exam revision or during periods of independent study. Thus, in Físchlár-Nursing, we have created scene- and topic-level keyframe browsing as well as a table of contents for each program in the Físchlár-Nursing system. These higher-level keyframes are semi-automatically created while the table of contents is manually created, and in combination they support rapid browsing and search through the collection, working at a level which is higher than shots and equivalent to news stories in Físchlár-News.

3 Experience with four video collections and the Físchlár system

The userbase of these four systems is quite large, with Físchlár-TV and Físchlár-News having a registered userbase of almost 2000 users, and Físchlár-Nursing will be used by several hundred student nurses on campus by the end of the present academic year. Físchlár-TREC2002 is a more specialist application and was used, in a controlled experiment, by only a few dozen users. The volume of video data accessible from the four Físchlár systems totals several hundred hours of video material.

The numbers of users involved and the volume of video material means that we are in a reasonably strong position to make statements about user information need types. What has made the development of these four separate video libraries interesting from a digital libraries point of view is the variety of user information need types and units of retrieval which users of the different systems have and which must be supported. This variety includes the granularity of determining what is "relevant" for a user (a shot, a news story, a coverage of a topic or an entire TV program or movie), as well as the balance between browsing vs. searching for video. Although all our Físchlár systems reported here share the same architecture and internal components such as indexing and analysis, all are very different in terms of interfaces and functionality.

The XML-based architecture of the system has been important for easy plug-in of different features catering to different applications under a single underlying platform. Most of the specific browsing features realised for different applications are the result of initial deployment followed by usage observation and refinement, and in technical terms their presence meant simple addition of Web requests by defining the required XML documents and corresponding XSL stylesheets to render this on the Web interface.

As our experience has shown, the differing requirements of different users of video IR systems create challenges for the designers of digital video libraries, and we have been genuinely surprised by this large variety in user requirements from different constituencies.

References

1. Browne P, Czirjek C, Gurrin C, Jarina R, Lee H, McDonald K, Smeaton AF, Ye J (2002) Dublin City University video track experiments for TREC 2002. In: Proceedings of the 11th text retrieval conference (TREC 2002), Gaithersburg, MD, 19–22 November 2002
2. Gaughan G, Smeaton AF, Gurrin C, Lee H, McDonald K (2003) Design, implementation and testing of an interactive video retrieval system. In: Proceedings of the ACM SIGMM international workshop on multimedia information retrieval (MIR 2003), Berkeley, CA, 7 November 2003
3. Jarina R, O'Connor N, Marlow S, Murphy N (2002) Rhythm detection for speech-music discrimination in MPEG compressed domain. In: Proceedings of DSP 2002 – 14th international conference on digital signal processing, Santorini, Greece, 1–3 July 2002
4. Lee H, Smeaton AF (2002a) Designing the user interface for the Físchlár digital video library. J Digital Inf 2(4)
5. Lee H, Smeaton AF (2002b) Searching the Físchlár-News archive on a mobile device. In: Proceedings of the 25th international ACM SIGIR workshop on mobile personal information retrieval, Tampere, Finland, 11–15 August 2002