Research in Information Managment at Dublin City University

Mark Roantree[1] Alan F. Smeaton[2]
[1]Interoperable Systems Group [2]Centre for Digital Video Processing
Dublin City University, Glasnevin, Dublin 9, IRELAND.
{mark, asmeaton}@Computing.DCU.ie

1 Introduction

The Information Management Group at Dublin City University has research themes such as digital multimedia, interoperable systems and database engineering. In the area of digital multimedia, a collaboration with our School of Electronic Engineering has formed the Centre for Digital Video Processing, a university designated research centre whose aim is to research, develop and evaluate content-based operations on digital video information. To achieve this goal, the range of expertise in this centre covers the complete gamut from image analysis and feature extraction through to video search engine technology and interfaces to video browsing. The Interoperable Systems Group has research interests in federated databases and interoperability, object modelling and database engineering. This report describes the research activities of the major groupings within the Information Management community in Dublin City University.

2 Multimedia Retrieval

The Centre for Digital Video Processing (CDVP) is a group of almost 25 researchers (Faculty, post-doctoral researchers and graduate students) which is funded by national and European research funding agencies as well as by the University and by industry. Since 1997 the CDVP has secured over €2m in research funds from these sources and has been part of 6 separate expressions of interest for research activities in the forthcoming EU Sixth Framework Programme which has positioned us well to extend our already large network of collaborators.

In order to make significant progress in indexing, browsing, linking and searching of large amounts of digital video information, skills and expertise in information management and in video coding and video analysis, are necessary. This is because so much analysis and processing of video is required in order to extract information from video on which to base content-based operations. The background to the CDVP is that it combines researchers with expertise in video coding and compression [2, 21], with those who have information retrieval and searching as the focus of their previous work [1, 7]. This has allowed us to generate a burst of activities in this area in recent years and to contribute to advancing the state of the art in video navigation. Since January 2000 we have had 35 refereed publications in areas such as video analysis, video indexing, video retrieval, video summarisation and video browsing, but in addition to producing research papers we are also practitioners, in that we concentrate a lot of our efforts into developing demonstrator systems. This helps to satisfy our funding bodies but also gives much focus and direction to our research as we regard these demonstrator systems as a crucial aspect of progressing our area of research. Before describing some of these search and browsing demonstrators, we will summarise our work on feature extraction from video

2.1 Feature Extraction from Video

In order to make any progress with extracting information from digital video information, the first task is to segment video into shots and select keyframes as summaries of each shot. A shot is defined as a single camera movement in time and, like many other groups, we have developed and tested techniques for shot boundary detection using colour histograms, colour moments, texture analysis, analysis

of shapes and even based on characteristics of the compressed video such as the distribution of B-frames in MPEG-1 [13]. Like other groups we have also evaluated the effectiveness of shot boundary detection, work in terms of recall and precision, though we evaluated our techniques on an 11-hour dataset, considerably larger than the size used by most other groups. Our greatest contribution in this area is the development of a shot bound detection technique which runs in the compressed domain, directly on MPEG-1 files, and the compute time for this is of the order of 5% to 6% of real time [5], whereas techniques which operate on the uncompressed domain normally run in about real time. This is a significant development.

Once shot boundaries are determined, video content is then subjected to a series of analyses in order to detect features. These can include keyframe selection, speech recognition (which we do not do), speech-music discrimination, speaker segmentation and identification, face detection, group detection, and camera motion detection [12, 8, 20]. In our work we do this in such a way that the output of these detectors, which forms the basis for searching and browsing of video archives, is marked-up in MPEG-7, allowing easy insertion into our video database systems.

Like most research groups working on advanced techniques for video navigation, we are working towards object-based segmentation, tracking, and identification directly from video sources, and on top of that we can then layer object-based searching, linking, browsing, etc. Towards that goal, one of our current projects involves performing such object segmentation and tracking, albeit in constrained domains initially, and using external metadata to further constrain the video analysis. Initial work on this is already showing progress.

2.2 Demonstration Systems

One of the strategic decisions made early in the lifetime of the CDVP is that demonstration systems are a good thing because they direct and provide focus for our research. This has helped to provide direction for our work on feature extraction from video and has allowed us to build and deploy the following systems:

- **Físchlár-TV** is a digital video recording, browsing, searching and playback tool which supports access to captured digital video content, currently taken from any of 8 free-to-air TV channels in Ireland [9]. Our users select programmes to be recorded from an online TV schedule which has genre (categories) assigned to each programme automatically, and as requested programmes are transmitted we capture the video, detect shots and keyframes and place the analysed programs into our video library. Users can then browse programme keyframes and playback content from any point in the recorded video. We also capture teletext, 24/7 from 6 TV stations and have developed a rudimentary text searching facility to allow video navigation as a compliment to keyframe browsing. Físchlár-TV is available throughout the University campus, library and residences and is used for research and for teaching and learning. It is extremely popular with over 1,800 registered users, about half of whom are "active" users, and during semester we record approximately 20 to 25 hours of content per day. At any point in time we can store up to 300 hours (400 programmes) online and our video server can support over 250 simultaneous playback streams (each MPEG-1 stream is 1.5 Mbps).

- **Físchlár-News** is essentially the same as Físchlár-TV except that the content is the main evening news from the national broadcasting TV station (RTÉ) which we record faithfully but in addition to supporting keyframe browsing of approximately 200 hours of TV news, we have a much more elaborate and useful text search facility which allows users to search for fine segments of TV news programmes [10]. Físchlár-News is deployed throughout our campus and is also available from University libraries of other Universities in Ireland. Our work in this area is now focussed on news story segmentation, identifying the boundaries between different news stories within a single broadcast using the closed caption text [19], the detection of the news anchorperson on screen and the story image projected behind them [12], speech-music identification [8], speaker discrimination, and advertisement detection [20]. It is by combining the respective outputs of these analyses carried out in parallel that we expect to achieve a high degree

of accuracy in story segmentation, and allow automatic linking of related stories thus allowing an even more useful system for video navigation to be developed.

- **Físchlár-Nursing** is another variation of the Físchlár system where the content consists of several hours of professionally produced video materials to be used as part of course delivery in our School of Nursing. In this system we hand-crafted a collection of indices on this video (in MPEG-7) on top of the automatic shot bound definition and keyframe selection, to allow students to be selective in what material they access. The system as a whole allows large-scale deployment of this video content to many hundreds of students on campus, and will also allow off-campus access to the various outreach centres throughout the country where the School of Nursing has teaching activities.

- **Físchlár-TREC2002** is the final variation of the Físchlár system which was developed specifically to allow our participation in the TREC video track in 2002. TREC is an annual bench-marking exercise for Information Retrieval (IR) applications, co-ordinated by the National Institute for Standards and Technology (NIST). Over the last decade it has spawned over a dozen "tracks" or specialist IR applications including retrieval from spoken audio documents, interactive IR and IR from web documents. In 2001 a track specialising in IR from video was introduced which attracted over a dozen participants. A corpus of almost 40 hours of video was distributed among participants, as was a collection of 25 user queries. Queries were run by each participant and each site's results for each query sent back to NIST, who pooled submitted results and performed manual relevance assessments on the pool. Once completed, this allows comparison among participating groups' results in terms of measures such as precision and recall. What was novel about the TREC2002 video track was that several sites ran feature detectors on the video search collection and made their outputs available to other groups. For example, we provided the outputs from our face detector, and our speech and music discriminator,

and used the speech recognised transcript from Microsoft Research Asia and the indoor/outdoor and landscape/cityscape detectors from IBM Research, all information marked up in MPEG-7. At the time of writing the performance of our submission in TREC2002 video track is unknown but the system we developed was an interactive search tool, allowing use of features and was combined with the usual Físchlár browsing capabilities.

These demonstration systems all have a web interface. Físchlár-TV has been in operation for almost 3 years, Físchlár-News is in operation for only 12 months, and Físchlár-Nursing and Físchlár-TREC2002 are recent developments.

2.3 Platforms and Interfaces

All of the demonstration systems mentioned above are XML-based and generate their outputs in XML which are transformed via XSL stylesheets into interfaces for users. The rationale for doing this is to allow easy porting of these systems to mobile platforms and to date we have successfully ported Físchlár-TV to the Compaq iPAQ running the Pocket-PC operating system and accessed over a wireless LAN. We also have a trial running with a local network provider to provide GPRS access via an xda mobile device.

While device-independence is an important aspect of our work from a technical perspective, the design of a consistent user interface which addresses both the user's needs and the user's constraints during use is even more important. We have developed a theoretical framework for video browsing from which we have been able to develop several video keyframe browsing implementations, each suited to the particular characteristics of a desktop PC or mobile platform [9]. This theoretical underpinning has served us well in helping us to develop our various interfaces while taking user needs into consideration.

3 Interoperable Systems and Data Engineering

The Interoperable Systems Group (ISG) has been funded seven times at national level to develop research projects in areas such as federated health-

care systems, distributed component technologies, object data modelling, multimedia databases, and mobile computing. The group currently comprises ten researchers who are involved in research initiatives across three separate projects. The IOMPAR project concerns the secure transfer of non textual data across a federated database architecture; the EGTV project is based on distributed multimedia database research; and the XPeer Project is a collaboration between the ISG and the University of Montpellier in France, for research into an XML federated database system.

3.1 Secure Transfer of Complex Objects

The IOMPAR Project (Internet Object Management: PRivacy Architecture) is focused on the secure transfer of complex objects [3]. It concentrates on limitations to existing e-commerce technologies: data residing in traditional relational databases, and security being difficult to guarantee across the network. The project assumes that data is not flat in structure, but takes the form of objects, which have a hierarchical structure and contain built-in operations. This implies that data retains the rich semantics contained at the modelling stage. As part of the project, an implementation that offers secure transfer of data between objects using existing internet technologies, is under development. The prototype will include sophisticated security primitives as part of the structured objects, unlike current technology which encrypts data at a very low-level thus, removing all structure.

IOMPAR makes use of some of the research output from its predecessor, the OASIS project [16], where a healthcare federation was constructed. A view mechanism for ODMG databases was constructed by extending the ODMG metamodel to accommodate virtual entities [15], and this is used to facilitate the restructuring of objects in views. This technology is reused to extract object views for encryption processes.

Research is focused on four main areas. The first is on a model transformation service for ODMG-to-XML data which requires the capture of all semantic information including behaviour. A second research task involves the specification and implementation of the Repository Query Language which provides a

reflection facility for ODMG and Object-Relational databases. This builds on previous work [14] in that it must provide an extended functionality and interface to operate in a mobile computing environment. A third task focuses on the specification of the prototype/clone transaction protocol for the transaction of objects of an unknown type. This work extends the basic ideas presented in [6]. A final task involves the implementation and integration of security primitives within the object structures. This task incorporates the implementation of smart-card based access control mechanisms to manage authorised access to sensitive object members, using state of the art encryption technology.

There are three industrial partners in the project, Versant Limited the makers of an ODMG database, Iona Technologies the makers of the Orbix distributed object software, and Orbism who act as the end-user for the prototype. A prototype system which offers the secure transfer of object schemas is due for completion in February 2003.

3.2 Multimedia Database Transactions

The aim of the EGTV (Efficient Global Transactions for Video) [17] research project is to provide query and update capabilities for a large distributed library of multimedia (video, audio and graphical) objects. Technical issues addressed are in the areas of storage, distribution, and global transactions.

The first task was to design the database metamodel for multimedia objects. Two database types are used for persistence: "pure" object-oriented databases and object-relational databases. Both offer identical interfaces to global query services. The two-pronged approach was aimed at exploiting developments in both the object-oriented and object-relational database areas. It also exploits previous research where MPEG objects have been processed and indexed to extract key frames at selected points in the MPEG file [11]. These frames are stored and used in later update and retrieval transactions. A full specification of the metamodel with mappings to both ODMG and Object-Relational databases is provided in [4].

Conventional storage technology either fails to provide for multimedia storage (relational or object-

relational technology), or adopts a purpose-built approach (Oracle Video Server, RealNetworks video streaming) that provides only a centralised storage capability with no fine-grained storage or retrieval techniques. An ongoing task is to build a distributed object-based storage architecture for multimedia objects. One of the key tasks to the success of this architecture is the development of a global catalog service, which provides smart metadata access across the federation.

Many of the scenarios for this project require that global views be created, and in some cases updated. This necessitates the creation of a global view service where multimedia objects from different sources can (conceptually) be placed side by side. It is intended that each multimedia server should designate those multimedia objects that may participate in global views. These objects may subsequently form part of many global views. This function will require a local view service for multimedia servers, and a global view service for remote users (such as users of hand held devices). The requirement for the placement of certain services (behaviour) in views provides a unique problem for the view mechanism.

It is not sufficient to adopt current transaction protocols due to the size of the objects exchanged. A transaction service is proposed which operates in two phases: a metadata phase and a data phase. The purpose of the metadata phase is to ensure that the correct multimedia object has been identified, and the correct time and segment information has been specified. The data phase then performs the physical data transaction. This requires the specification of a new query language, and it is intended that this language be XML-based to exploit future developments in popular technology standards.

3.3 XML Databases

The aim of the XPeer project is to specify all areas of research and development of a distributed XML database, including architecture, software components, query language, and the mapping of XML and object-oriented technologies. The project is currently in the process of designing a distributed XML database architecture using peer-to-peer technology. Future research is plannned on object database query languages, global data dictionaries, an XML Database Interface, and distributed XML transactions. The project began in 2001 and is due to complete in 2003.

4 Conclusions

All of the activities described in this report are ongoing. The Information Management research group maintains its main webpage at `www.compapp.dcu.ie/research/information` with an outline of major activites, while the individual research groups such as the Centre for Digital Video Processing (`www.cdvp.dcu.ie`) and the Interoperable Systems Group (`www.compapp.dcu.ie/~isg`) maintain their own sites with project details and online publications and software demonstrations.

References

[1] Agosti, M. and Smeaton, A.F. (Eds). *Information Retrieval and Hypertext*. Kluwer Academic Publishers, 1996.

[2] Brady, N. and O'Connor, N. Object Detection and Tracking using an EM-based Motion Estimation and Segmentation Framework. *Proceedings of the IEEE International Conference on Image Processing* (ICIP'96), Vol. 1, Lausanne, Switzerland, September, 1996.

[3] Bećarević D. and Roantree M. Distributed Transactions for ODMG Federated Databases. In *New Developments in Distributed Applications and Interoperable Systems* (DAIS), Zielinski et al (eds), pp. 317-322, Kluwer Academic Press, 2001.

[4] Bećarević D., Subieta K., and Roantree M. A Metadata Approach to Multimedia Database Federations. *Submitted for publication*, 2002.

[5] Calic, J., Sav, S., Izquierdo, E., Marlow, S., Murphy, N. and O'Connor, N. Temporal Video Segmentation for Real-Time Key Frame Extraction. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (ICASSP 2002), Orlando, USA, 17-23 May, 2002.

[6] Hasselbring W. and Roantree M. A Generic Framework for Database Interoperability. Proceedings of 3rd IFCIS International Conference on Cooperative Information Systems (CoopIs98), pp. 64-73, IEEE Computer Society Press, 1998.

[7] Harman, D., Schauble, P. and Smeaton, A.F. Document Retrieval. In *Survey of the State of the Art in Human Language Technology, Linguistica Computazionale*, Vol XII-XIII, 1997.

[8] Jarina, R., O'Connor, N., Marlow, S. and Murphy, N. Rhythm detection for speech-music discrimination in MPEG compressed domain. *Proceedings of the 14th International Conference on Digital Signal Processing* (DSP 2002), Santorini, Greece, 1-3 July, 2002.

[9] Lee, H. and Smeaton, A.F. Designing the user-interface for the Físchlár Digital Video Library. *Journal of Digital Information*, 2(4), 2002.

[10] Lee, H. and Smeaton, A.F. Searching the Físchlár-NEWS Archive on a mobile device. *Proceedings of the 25th International ACM Conference on Research and Development in Information Retrieval* (SIGIR 2002), Workshop on Mobile Personal Information Retrieval, Tampere, Finland, 11-15 August, 2002.

[11] Lee H., Smeaton A., O'Toole C., Murphy N., Marlow S. and O'Connor N. The Físchlár Digital Video Recording, Analysis and Browsing System. *Proceedings of the RIAO 2000 Content-based Multimedia Information Access*, 2000.

[12] O'Connor, N., Czirjek, C., Deasy, S., Marlow, S., Murphy, N. and Smeaton, A.F. News story segmentation in the Físchlár Video Indexing System. *Proceedings of the 8th International Conference on Image Processing* (ICIP 2001), Thessaloniki, Greece, 7-10 October, 2001.

[13] O'Toole, C., Smeaton, A.F., Murphy, N. and Marlow, S. Evaluation of automatic shot boundary detection on a large video test suite. *Proceedings of the 2nd Conference on Image Retrieval: The Challenge of Image Retrieval* (CIR '99), Newcastle, U.K., 25-26 February, 1999.

[14] Roantree M., Kennedy J., and Barclay P., Using a Metadata Software Layer in Information Systems Integration, *3th Conference on Advanced Information Systems Engineering (CaiSE)*, pp 299-314, LNCS 2068, Springer, 2001.

[15] Roantree M., Kennedy J., and Barclay P., Constructing View Schemata Using an Extended ODL, *9th International IFCIS Conference on Cooperative Information Systems (CoopIS)*, pp 150-162, LNCS 2172, Springer, 2001.

[16] Roantree M., Murphy J., and Hasselbring W. The OASIS Multidatabase Prototype. *ACM SIGMOD Record*, 28(1), 1999.

[17] Roantree M. Efficient Global Transactions for Video Media. *Technical Report No. ISG-02-03*, Dublin City University, 2002.

[18] Roantree M. Metadata Management in Federated Multimedia Systems. *Australian Computer Science Communications 24:2*, pp. 147-155, 2002.

[19] Stokes, N., Carthy, J., & Smeaton, A.F. Segmenting broadcast news streams using lexical chains. *Starting Artificial Intelligence Researchers Symposium* (STAIRS 2002), Lyon, France, 22-23 July, 2002.

[20] Sadlier, D., Marlow, S, O'Connor, N. and Murphy, N. Automatic TV advertisement detection from MPEG bitstream. Proceedings of the 3rd International Conference on Enterprise Information Systems, *Workshop on Pattern Recognition in Information Systems* (WPRIS 2001), Setubal, Portugal, 7-10 July, 2001.

[21] Salembier, P., O'Connor, N., Correira, P. and Pereira, F. Hierarchical Visual description schemes for still images and video sequences. *Proceedings of the IEEE International Conference on Image Processing* (ICIP'99), Kobe, Japan, October, 1999.