Adaptive Information Cluster

Dublin City University

Hyowon Lee
Adaptive Information Cluster,
Dublin City University, Glasnevin,
Dublin 9
Email: hlee@computing.dcu.ie

Sinéad Mc Givney
Adaptive Information Cluster,
Dublin City University, Glasnevin,
Dublin 9
Email: smcgivney@computing.dcu.ie

Daragh Byrne
Adaptive Information Cluster,
Dublin City University, Glasnevin,
Dublin 9
Email: dbyrne@computing.dcu.ie

Alan F. Smeaton
Adaptive Information Cluster,
Dublin City University, Glasnevin,
Dublin 9
Email: asmeaton@computing.dcu.ie

ABSTRACT
The Adaptive Information Cluster (AIC) is a collaboration between Dublin City University and University College Dublin, and in the AIC at DCU, we investigate and develop as one stream of our research activities, various content analysis tools that can automatically index and structure video information. This includes movies or CCTV footage and the motivation is to support useful searching and browsing features for the envisaged end-users of such systems. We bring in the HCI perspective to this highly-technically-oriented research by brainstorming, generating scenarios, sketching and prototyping the user-interfaces to the resulting video retrieval systems we develop, and we conduct usability studies to better understand the usage and opinions of such systems so as to guide the future direction of our technological research.

Author Keywords
Interaction Design for Video Retrieval, HCI, Content-Based Video Analysis, Collaborative Interface, User Study, Usage Monitoring, Passive Capture

BACKGROUND AND DESCRIPTION
The Adaptive Information Cluster (AIC) is a cross-disciplinary research cluster that brings together cognitive and computer sciences, software system engineering, electronic engineering and chemistry, to integrate research on adaptive sensor networks, content extraction, adaptive utilization and adaptive middleware. The AIC was established in 2004 with €5.6 Million in funding from Science Foundation Ireland and has subsequently more than doubled that funding, with the current size of about 90 postgraduate students and post-doctoral researchers across the two campuses.

The Cluster's content-based image/video analysis research is based in the Centre for Digital Video Processing at Dublin City University, and investigates various ways to automatically index visual materials such as personal photos, surveillance video footage, TV programmes and movies, in order to allow an effective and easy browsing and searching for end-users. The application areas are very broad in the sense that any domain that needs to manage large amount of video materials could benefit automatic indexing and retrieval, and in particular CCTV-type applications will allow automatic content extraction and analysis of the video-captured environment, augmenting other 'sensing' dimensions within the Cluster. For example, research on molecular sensors that can adapt its selectivity and sensitivity to its environment is carried out in the Adaptive Sensors Group within the Cluster (see the linked presentation by Sarah Brady et al. for more detail on the molecular sensing aspects of the AIC research programme).

The automatic image/video indexing techniques we research include shot/scene/event boundary detection in video, face/object detection and tracking, detection of indoor/outdoor and landscape/cityscape settings, fusion of multiple evidences from feature extraction in video, automatic sports summarization, and augmenting content-based analysis with context information such as time/location. Potential applications of these automatic analyses in the real world is huge. For example, with the ubiquitous digital camera and camera phones today, the

number of digital photos that people take has been dramatically increasing, and the popularity of community-based online services such as Flickr and YouTube is becoming greater. Although an owner of photos is expected to add captions or annotations for each photo in order to help search/browse effectively later on, manual annotation is a burden for the user which is getting heavier, in proportion to the increasing quantity of owned photos. With automatic content analysis tools, photos/videos can be automatically annotated, organised, and structured without any human effort (although the users should be allowed to manually add more annotations if wished). Imagine a Flickr-like service where a user uploads a large number of photos but without having to spend time and effort to individually annotate each photo.

Similarly, the amount of TV programmes broadcast daily by many broadcasting companies is so large that it is impossible to get human indexers to describe for each shot/scene, the names of people/objects appearing in it, and many other visual and conceptual aspects of the contents of the video for every programme. Automatic content analysis can be an extremely valuable tool to enable such detailed descriptions without the expensive human involvement.

Our focus on the content analysis of image/video is to develop these tools in order to support subsequent access by users without putting enormous effort in manual indexing/annotation.

RESEARCH AGENDA

While technological research in developing automatic tools for image/video indexing/retrieval remains our main agenda, we also investigate aspects where the possible end-users interact with such tools. Technology-driven research is necessary, but should always be checked and guided by how the outcomes of the research could and should be applied in real-life situations.

For example, our content analysis to automatically detect and track a person in video is an on-going research topic within our centre. However, exactly where and how such a technology, if matured, could be applied to support a specific task for a specific group of users in a specific context is an important issue to be address in parallel with the underlying technology development itself. As a parallel stream of the study of person detection/tracking algorithms, we investigate interaction schemes where a user can view automatically detected persons in video, select/unselect them, and use them for subsequent request to the system for querying. Without the interaction design for these, the detection algorithm itself will remain only a *potential* technology, however accurate and efficient it is.

Thus, for each of the main projects we have, we have been allocating a sub-project investigating the user/usability and interaction design issues associated with the developing technology in that project. This starts with active brainstorming, scenario development, iterative sketches and prototyping, followed by various forms of user experiments with the prototype or working systems.

For example, since 2001 we have been developing an interactive video retrieval system each year in our participation to the annual TRECVid digital video evaluation activity (http://www-nlpir.nist.gov/projects/t01v/t01v.html) organised by the National Institute of Standards and Technology (NIST) in the U.S. Typically a system we develop contains around 100 hours of news video data, and users can query, browse, and play clips to find a particular video shot. We conduct user experiments in which we invite 12-30 users and ask them to engage in a set of search tasks using the system, from which we obtain their performance, usage data, and subjective satisfaction data. In 2005 the system we developed and conducted a user study on was a novel interface where two users collaboratively searched for video clips on a touch-sensitive, horizontal tabletop interface (described in more detail in the Sampling of Research Projects section). From these series of user experiments on novel, experimental video retrieval systems over the years, we have gained valuable experience and insights into the design/user/usability aspects of such systems. This, in turn, gets fed back to the technological research direction and sets the direction to be more tuned to real users' perspectives.

Another example is one of our long-term projects to develop an automatic TV news analysis and delivery system on the web, called Físchlár-News. Everyday the system recorded the 9 o'clock RTE 1 TV news and automatically structured recorded videos into shots and news stories, established links among similar news stories from different dates, and provided a web-based interface where a user could search, browse, watch and get recommendations of the news story videos. While the underlying technology has been the main subject of the project (including news story segmentation, anchorperson detection, advertisement detection, etc.), the system's web-based interface has been a separate stream of research in itself. By having a robust, web-based system deployed on a campus, we turned the whole university campus into our usability laboratory to monitor people's daily usage, access patterns over time, and their reactions, opinions and complaints after having used the system for a long period of time.

As seen in the above examples, in our centre the technical research goes hand-in-hand with more user-oriented, HCI perspective research that leverages its technical research output.

EDUCATIONAL EFFORT

The AIC currently has about 60 full-time postgraduate students at Masters and PhD level engaged in various aspects of the AIC research. In addition, we create smaller size projects with more in-depth investigation and we bring in 3rd and 4th year undergraduate students and hire summer

interns to engage in the specific areas of the research, providing a mechanism for knowledge transfer.

FUTURE PLANS/DEVELOPMENTS

The AIC has grown much since its start nearly 4 years ago in the scope of research, in the amount of industry funding obtained and in the scale of the network of educational and industrial connections it fostered.

In terms of content analysis of image/video, we are moving from low-level syntactic analysis such as shot detection, colour/shape/texture analysis, etc. to high-level semantic analysis such as person present, action/dialogue sequences, meaningful event grouping, pedestrian detection in CCTV, assigning label/name for detected person/object, etc. The current status of such semantic analysis is still at an early stage of research, and we are trying to enhance the research community's rather slow progress by leveraging context information obtained at the time of image/video capture. Regarding interaction design and usability studies, we are developing more fundamental and theoretical foundations for user-interfaces to automatic multimedia indexing systems and at the same time working on developing more comprehensive and generic user evaluation strategies for such systems.

A SAMPLING OF RESEARCH PROJECTS CONDUCTED IN THE GROUP

Project: Personality and Interface Matching for Multi-User Collaborative Applications

To date, much development has tasken place in the area of Groupware or Computer Supported Cooperative Work (CSCW). However, research that focuses on the end-users in the area of multi-user, collaborative, tabletop technologies is limited. In this project we investigate this gap between technological development and user consideration in a collaborative setting. We aim to provide a set of interface guidelines that give maximum support to users of an application developed for such technologies, and to propose a methodology that facilitates matching certain system interface variants to a pair of users of a multi-user, collaborative, tabletop device. These matching guidelines are derived from information mined from the users' personality information, interaction styles and performance results on two or more variants of interfaces to particular genres of collaborative applications. (Duration: 2004 – 2007; Partners: Mitsubishi Electric Research Lab (MERL); Funding agencies: SFI, IRCSET, MERL)

Project: Visualising a SenseCam LifeLog Diary

SenseCam, developed by Microsoft Research, is a passive photo capture device with a number of in-built sensors. Worn around a user's neck, the device takes photos automatically triggered by its sensors' change detection. With around 3,000 photos automatically taken each day, access to the photos is hindered by the sheer amount of photos especially when the user has been wearing the device for days, weeks, months or even years. Leveraging automatic content analysis techniques within our centre, we study how the photos could be automatically organised into coherent events, how the novelty or importance of each event could be automatically established, and how these could be presented as an interactive diary interface for the user to easily search and browse tens of thousands of photos. Bringing in our experience in interaction design for multimedia retrieval systems, we are developing interactive online visual diaries that allow an effective and engaging SenseCam photo searching/browsing experience for the user. (Duration: 2005 – 2008; Funding agencies: SFI, Microsoft Research)

Project: MovieBrowser

In this project we develop the tools which automatically detect different types of scenes found in feature films. By combined analysis of speech/music, degree of motion, duration of shots, and speaker change rates in the groups of camera shots in movies, the system can automatically select specific types of movie scenes, namely "exciting scenes", "dialogue scenes", and "musical scenes". Dozens of movies were analysed and an integrated interface that allows a user to browse those types of scenes in a convenient way were designed. By being able to quickly filter out a large number of scenes in a 2-3 hour movie in order to find out a particular scene, the system enhances the conventional "scene access" features of commercial DVD menu systems, in a fully automatic way. We deployed this system in the computer labs in the School of Communications within DCU by installing the necessary software and getting it incorporated as the teaching/learning tool for one of the students' modules in Film Studies. From the students' point of view, being able to watch the syllabus' movies in their computer lab is in itself a great convenience, and the main motivation for using the system. We are currently conducting interviews and focus group analyses to ascertain usability problems, to refine its interface and to incorporate new features. (Duration: 2002 – 2007; Funding agency: Enterprise Ireland, Malaysian Government, SFI)

RESOURCES FOR FURTHER READING

Adaptive Information Cluster website:

http://www.adaptiveinformation.ie/

Centre for Digital Video Processing website:

http://www.cdvp.dcu.ie/

SenseCam Project at the AIC:

http://www.cdvp.dcu.ie/SenseCam/

OTHER ACTIVITIES

The AIC is involved in a number of research communities, consortia, and collaborative projects.

K-Space - Knowledge Space of Semantic Inference for Automatic Annotation and Retrieval of Multimedia Content (EU FP6 Network of Excellence)

aceMedia - Integrating knowledge, semantics and content for user-centred intelligent media services (EU FP6 Integrated Project)

MediAssit - Tools for organizing, browsing and retrieving from a personal electronic picture collection.

International events held:

CIVR 2004 (International Conference on Image and Video Retrieval), Dublin City University, 21-24 July 2004.

ESSIR 2005 (European Summer School in Information Retrieval), Dublin City University, 5-9 September 2005.

CONTACT POINTS

Details of contact person(s) for the group/centre

Prof. Alan F. Smeaton

The Adaptive Information Cluster

Dublin City University

Glasnevin, Dublin 9

Tel: +353 -1 700 5262

Email: asmeaton@computing.dcu.ie

SAMPLE PUBLICATIONS

Collaborative Video Searching on a Tabletop. Smeaton A.F, Lee H, Foley C and Mc Givney S. *Multimedia Systems Journal*, 12(4-5), 2006.

Interactive Searching and Browsing of Video Archives: Using Text and Using Image Matching. Smeaton A.F, Gurrin C and Lee H. *In: Hammoud, Riad (Ed.), Interactive Video: Algorithms and Technologies, 2006, XVI, 250 p. 109 illus., Hardcover, ISBN: 3-540-33214-6*, 2006.

User Evaluation of Físchlár-News: An Automatic Broadcast News Delivery System. Lee H, Smeaton A.F, O'Connor N and Smyth B. *TOIS - ACM Transactions on Information Systems*, 24(2), 2006.

Multimodal Segmentation of Lifelog Data. Doherty A, Smeaton A.F, Lee K Ellis D. *RIAO 2007 - Large-Scale Semantic Access to Content (Text, Image, Video and Sound)*, Pittsburgh, PA, USA, 30 May - 1 June 2007.

Event Detection in an Audio-based Sensor Network. McHugh M and Smeaton A.F. *Multimedia Systems Journal*, 12(3), 2006.

Techniques Used and Open Challenges to the Analysis, Indexing and Retrieval of Digital Video. Smeaton A.F. *Information Systems Journal*, 32(4), 2006.