

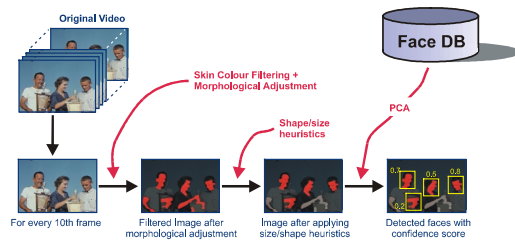
P. Browne, C. Czirjek, C. Gurrin, R. Jarina, H. Lee, S. Marlow, K. Mc Donald, N. Murphy, N. O'Connor, A. F. Smeaton, J. Ye

Dublin City University's Centre for Digital Video Processing (CDVP) participated in the Feature Extraction and Search Tasks. In the Feature Extraction Task, we submitted results for Face, Speech, and Instrumental Sound Features. In the Search Task, we used a tailored version of the Físchlár Digital Video System which allowed video query and browsing based on all 10 Features (some donated by IBM and MSRA) and an ASR Transcript (donated by LIMSI). Our interactive search experiment was conducted with 12 users for all 25 topics.

Feature Extraction Task (Face, Speech, Instrumental Sound)

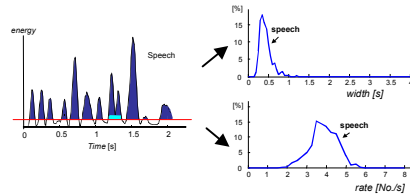
Face Extraction

- Every 10th frame used for processing
- Skin colour filtering followed by cleaning-up with morphological adjustment & size/shape heuristics
- Principal Component Analysis with an example face database



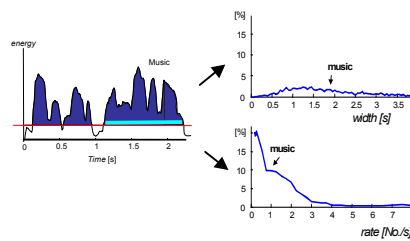
Speech Extraction

- Short width of energy peaks (<1s)
- Rates of energy peaks are concentrated around 4/s
- Necessary data is taken directly from MPEG-1 audio bitstream (Scalefactors)

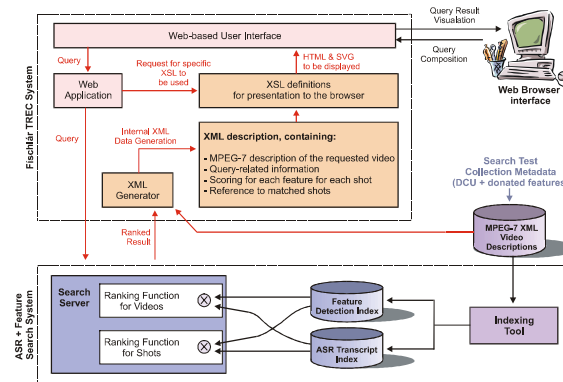


Instrumental Sound (Music) Extraction

- Wide width of energy peaks (>4s)
- Infrequent occurrence of peaks (e.g. 1/4s)
- Existence of rhythm
- Existence of harmonicity
- Necessary data is taken directly from MPEG-1 audio bitstream (Scalefactors)



Search Task (Interactive)



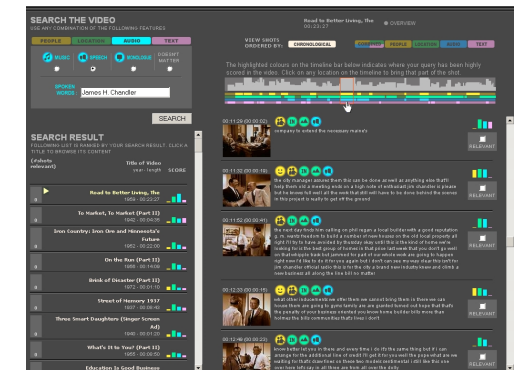
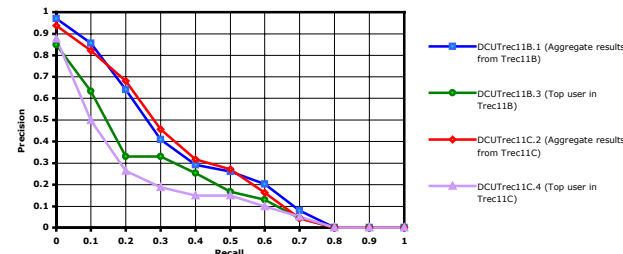
Interactive searching & browsing system

- Físchlár Digital Video System tailored for the Search task
- Contains 40 hours of search test collection
- 10 Features + ASR Transcript query interface
- XML-based architecture with MPEG-7 video descriptions
- XSL stylesheets render HTML and an SVG timeline for a conventional web browser
- Shot list presents a keyframe, detected Features, ASR transcript, and query score bars for each shot in a specific video

The Interactive Search Experiment

Our objective was to evaluate the benefit of incorporating features into interactive video retrieval.

- Two different retrieval systems were compared:
 - Features + ASR Transcript searching (Trec11B)
 - ASR Transcript-only searching (Trec11C)
- 12 users, each conducting all 25 topics
- Each topic was evaluated 12 times, 6 times for each interface giving a total of 300 query evaluations
- Users were under a time-constraint of 4 minutes maximum allowed per topic
- Coffee, cookies and headphones were provided !



Interface for Features + ASR Transcript searching (Trec11B)