The (IM)2 Newsletter
The Multimodal Media File Server is online, the Quarterly IP Status Reports

The Computer Vision at IDIAP
studies problems in machine visual perception

Issue N° 11
February 2003
Editor: Jean-Albert Ferrez
www.im2.ch/nl nl@im2.ch

INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

The (IM)2 Newsletter

Every month the (IM)2 Newsletter brings you the latest and hottest scientific and administrative news about the (IM)2 NCCR and related topics

The Multimodal Media File Server is online

mmm.idiap.ch

The Multimodal Media File Server is a repository for storing audio and video recordings to support research on multimodal information processing. Its purpose is to allow partners in the (IM)2 and M4 research projects to begin sharing raw and processed audio and video data with each other. The initial data available is a set of meeting recordings taken from the IDIAP smart meeting room. Each recording consists of output from three cameras at full PAL resolution and frame rate, plus audio from at least one microphone array and lapel microphones for each participant. Additional recordings from other sources (e.g. ICSI and Fribourg) will also be made available from this server in the future.

The server provides HTTP, RTSP, and FTP interfaces to support browsing, playing, retrieving, and adding of recorded multimodal data files. It can also serve as a platform to support future browsing and searching applications. The server is hosted at IDIAP and was developed in collaboration with members of (IM)2.AP, (IM)2.DS and (IM)2.MDM.

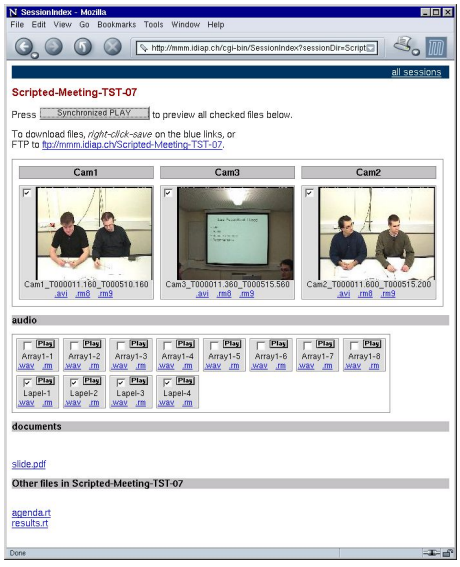

Figure 1: Browsing a session

BROWSING: Each recorded session has a directory on the file server and a dynamically generated "home page" that displays all available files including a jpeg image for each video file [Figure 1]. Every file is downloadable from this page by FTP or HTTP.

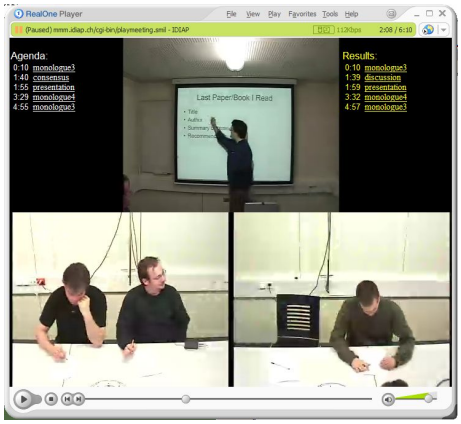

Figure 2: Playing several streams from one session

PLAYING: The session "home page" has images and buttons to stream any audio or video file using RealPlayer on Unix or Windows, and a "Synchronized Play" button that dynamically generates a SMIL presentation from all user-selected "checked" media clips to display them simultaneously in sync [Figure 2]. Start time offsets and durations are part of the file names to allow the media file server (and other software) to account for varying start times of concurrently recorded audio and video files.

RETRIEVING: Data can be downloaded for processing on your local computer using either HTTP or FTP. To download over HTTP, right-click-save on the desired filename extension in your web browser. To use anonymous FTP, simply connect to ftp://mmm.idiap.ch/ with your favorite FTP client.

ADDING: Tools are available to rename and format recorded data so it can be displayed, previewed, and retrieved using the web and SMIL user interfaces on this server. At this time, however, new data files are still added manually by server administrators. If you have multimodal media data to contribute, please contact us at mmmAdmin@idiap.ch and we will be happy to work with you to make the data available on this server.

Quarterly IP Status Reports

The Quarterly IP Status Reports for the period October - December 2002 are available from the local (IM)2 web pages. These internal reports list the achievements of the individual projects during the period under review and provide links to further material such as web pages and publications. They are a good way to keep up with the activity of other IPs.

Events

(IM)2.SA workshop **5.2.03**

The teams involved in (IM)2.SA will meet for an intensive one-day workshop early February at EPFL. Achievements in the 4 core projects and the 5 related White Papers will be presented and discussed.

(IM)2 Scientific and Industrial Advisory Boards **13–14.2.03**

The first joint meeting of the (IM)2 Scientific and Industrial Advisory Boards will take place in Martigny on February 13 and 14. The two day meeting will start with a general presentation of the NCCR and its objectives, and then focus on three topics: input modalities, multimodal processing, and applications.

ICASSP Special Session on Smart Meeting Rooms **6–10.4.03**

IDIAP will chair a Special Session on Smart Meeting Rooms at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), in Hong Kong, April 2003. The session consists of invited papers by leading research groups, who will present current work on various aspects of this emerging domain. Details at www.icassp2003.com.

Eurospeech 2003 **1–4.09.03**

IDIAP is organizing the next Eurospeech'2003 international conference, which will be held at the International Congress Centre of Geneva in September 2003. Eurospeech is the premiere conference on speech and language technology, attracting more than 1000 scientists every two years. Details at www.eurospeech2003.org.

The (IM)2 Newsletter is a publication of the NCCR on Interactive Multimodal Information Management, hosted by the Dalle Molle Institute for Perceptual Artificial Intelligence, Martigny, Switzerland. The National Centers of Competence in Research are managed by the Swiss National Science Foundation on behalf of the Federal Authorities.

SCHWEIZERISCHER NATIONALFONDS
FONDS NATIONAL SUISSE
SWISS NATIONAL SCIENCE FOUNDATION

The (IM)2 Newsletter backpage
Every month the (IM)2 Newsletter backpage presents one research group involved in the (IM)2 network: people, projects, research themes, infrastructure, etc.

Issue N° 11
February 2003
Editor: Jean-Albert Ferrez
www.im2.ch/nl nl@im2.ch

INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

The Computer Vision Group at IDIAP

The computer vision group at IDIAP studies problems in machine visual perception, such as media annotation, people detection and human gesture tracking and recognition. Research activities center on analysis of visual and multimedia data, and improvement of basic detection and classification measures and algorithms. This improvement may be achieved by enhancing and extending existing algorithms, or by creating new algorithms and measures. This frequently involves collaboration with the two other groups at IDIAP, speech processing and machine learning, as complementary expertise is brought to bear on a problem.

There is strong expertise within the vision group in areas of text processing from both documents and video, object tracking and recognition of gesture, and domain-based video analysis. The group is active in all of these areas under a number of collaborative European and Swiss national projects.

People

The group is jointly led by three seniors, Dr Sébastien Marcel, Dr Jean-Marc Odobez and Dr Daniel Gatica-Perez, and is composed of ten PhD students, one engineer and one internship student.

The computer vision group at IDIAP. From left to right, back: Yann Rodriguez, Agnès Just, Mark Barnard, Florent Monay, Pedro Quelhas, Maël Guillemot, Frédéric Kottelat, Jean-Marc Odobez, Beat Fasel, Daniel Gatica-Perez; front: Fabien Cardinaux, Alessandro Vinciarelli, Sébastien Marcel, Kevin Smith, Silèye Ba.

Research Themes

Face Algorithms: Face algorithms can be divided into four different areas.

- Face detection: The goal of face detection is to identify and locate human faces in images at different positions, scales, orientations and lighting conditions.
- Face localization: Face localization is a simplified face detection problem with the assumption that the image contains only one face.
- Face verification: Face verification is concerned with validating a claimed identity based on the image of its face, and either accepting or rejecting the identity claim.
- Face recognition: The goal of face recognition is to identify a person based on the image of its face. This face image has to be compared with all registered persons. Therefore, face recognition is computationally expensive with respect to the number of registered persons.

The vision group is mainly interested in face detection and verification using neural networks, SVM based methods or boosted weak classifiers.

Gesture Recognition: Gestural interaction based on the image is the most natural method for the construction of advanced man-machine interfaces. Thus, machines would be easier to use by associating the gestural command with the vocal command. This includes recognition of gestures such as facial expressions, hand postures, hand gestures and body postures. Current work on facial expression recognition is based on convolutional neural networks. Statistical approaches (skin color blobs) for object segmentation (faces and hands) in color images are investigated. The vision group is also interested in gesture recognition using hybrid models (Hidden Markov Models and Neural Networks) such as Input/Output Hidden Markov Models.

Tracking and activity recognition: Object tracking represents an essential component of gesture recognition, human behavior monitoring, and video indexing. IDIAP is investigating the design of stable trackers that are robust against ambiguities, image measurements, changes in the acquisition setting, and object intra-class variability. The group focuses on two areas: (1) the development of sequential Monte Carlo (SMC) techniques, and the exploration of the combination of SMC and finite state motion models based on HMMs for joint tracking and recognition of people activity, and (2) the fusion of multiple visual and multimodal (audio-visual) features, for example for speaker tracking.

Multimedia content analysis: The vision group is developing statistical models, algorithms and tools to automatically extract relevant information from audio-visual streams, which can be used for structuring, annotating, indexing and retrieving multimedia databases. Some of the current research directions include:

- Media structuring: The structure of videos is needed both at the individual and at the database levels. On one hand, finding structure in individual videos (shots, scenes) is useful to generate video summaries for browsing and retrieval, and usually constitutes the starting point to extract higher-level information. On the other hand, structuring a whole video database is useful for access and filtering (locating video replicas, organizing by "visual topic", etc.).
- Event classification: The group is developing audio-visual feature extraction and data fusion algorithms for event classification in sports video and meeting databases. Current efforts have been directed to define semantically meaningful events, and to learn their statistical models for classification.
- Text Detection and Recognition in Images and Videos: The vision group is involved in text detection and segmentation algorithms, and also examination of new paradigms in video text recognition. The goal of current research is to exploit the temporal redundancy to fuse recognition results of the same text obtained at different times.
- Modeling of textual and visual features: Members of the group investigate joint statistical models of words and visual features in multimedia databases, to relate low-level visual information with semantics. Such models would allow for important information retrieval functionalities, like clustering (grouping images that refer to the same text topics), annotation (attaching words to visual content), and illustration (attaching images to words).

Handwriting recognition: Offline handwriting recognition is the automatic transcription of handwritten data when only its image is available. The group has developed a recognizer based on continuous density HMMs which can deal with single words as well as handwritten texts (with the help of Statistical Language Models).

sm-jmo-dgp

The (IM)2 Newsletter is a publication of the NCCR on Interactive Multimodal Information Management, hosted by the Dalle Molle Institute for Perceptual Artificial Intelligence, Martigny, Switzerland. The National Centers of Competence in Research are managed by the Swiss National Science Foundation on behalf of the Federal Authorities.

SCHWEIZERISCHER NATIONALFONDS
FONDS NATIONAL SUISSE
SWISS NATIONAL SCIENCE FOUNDATION