

L²F – Spoken Language Systems Laboratory The VIDI-Video European Project



Video is vital to society and economy. It plays a key role in the news, cultural heritage documentaries and surveillance, ant it will soon be the natural form of communication for the Internet and mobile phones. Hence the need for video search engines that can cope with the peta bytes of future video archives.

Search concepts are categorized as types of scenes, types of objects, people, and events. Each concept can be specific: *Cavaco Silva*, or generic: *a happy person*. Some concepts will imply an audio analysis (e.g. *a cheering audience*), visual analysis (e.g. *an office scene*), style analysis (e.g. *a monologue*), or multimodal analysis (e.g. *a goal in soccer*).

VIDI-Video

Interactive semantic video search with a large thesaurus of machine-learned audio-visual concepts

Goal. Boosting the performance of video search engines by forming a 1000 element *thesaurus* detecting instances of audio, video or mixed-media content.

Description. The project will apply machine learning techniques to learn many different detectors from examples, using active one-class classifiers to minimize the need for annotated examples. The project approach is to let the system learn many, possibly weaker detectors describing different aspects of the video content instead of modeling a few of them carefully. The combination of many detectors will render a much richer basis for the semantics. The integration of audio and video analysis is essential for many types of search concepts.

Application scenarios:

- broadcast news
- cultural heritage
- surveillance

Duration: February 2007 - January 2010

www.inesc-id.pt



researchers

L²F: The European Project VIDI-Video

UvA - Universiteit van Amsterdam, the Netherlands (coordinator) **Partners** UNIS - University of Surrey, UK UNIFI – Universita degli Studi di Firenze, Italy INESC-ID – Instituto de Engenharia de Sistemas e Computadores . Investigação e Desenvolvimento em Lisboa, Portugal CERTH – Centre for Research and Technology Hellas, Greece CVC - Centroi de Vision por Computador, Spain B&G – Stichting Netherlands Instituut voor Beeld & Geluid, the Netherlands FRD - Fondazione Rinascimento Digitale, Italy Subcontracting UoM - University of Modena e Reggio Emília, Italy IIT - Indian Institute of Technology, India **WorkPackages** Learning system WP4 Visual Analysis WP2 Video Processing _ . . (2b) **2**c WP3 Audio Analysis 2d WP8 Data and queries WP9 Dissemination WP5 Learning integrated feature detectors (1) \bigcirc 4 Run time system 6 WP6 Software WP7 Demonstrators development and applications 5 WP1 management 13 **INESC-ID** Task 2.3 Audio Segmentation Segmentation of the audio stream into homogeneous regions according to **Participation** background conditions and speakers. Task 3.1 Detection of Audio Events Perform audio event detection by machine learning techniques(goals in sport matches, shooting, explosions, car or helicopter noises, cries, screams, laughter, cheering, ... Task 3.2 Speech Recognition Build a large vocabulary continuous speech recognition system tailored to broadcast news **INESC-ID** Isabel Trancoso Main João Paulo Neto