

# Workshop on Very-Large-Scale Multimedia Corpus, Mining and Retrieval

<http://vls-mcmr10.eurecom.fr>

29 October 2010,

held in conjunction with

ACM International Conference on Multimedia  
Firenze, Italy



Pivotal to many tasks in relation to multimedia research and development is the availability of sufficiently large datasets and their corresponding ground truth as well as the tools for efficiently handling and managing such large datasets. Currently available datasets for multimedia research are either too small such as the Corel or Pascal datasets, too specific like the TRECVID dataset, or without ground truth, such as the several recent efforts by MIT and MSRA that gathered millions of Web images for testing. While it is relatively easy to crawl and store a huge amount of data, the creation of ground-truth necessary to systematically train, test, evaluate and compare the performance of various algorithms and systems is a major problem. For this reason, more and more research groups are individually putting efforts into the creation of such corpus in order to carry out research on large-scale dataset. There is a need to unify these individual efforts into the creation of a unified web-scale repository which would benefit the entire multimedia research community.

The purpose of this workshop is to bring together researchers interested in the construction and analysis of Web Scale multimedia corpus and methodologies to mine and retrieve from them. The Workshop will provide a forum to consolidate key factors related to research on very large scale multimedia dataset such as the construction of dataset, creation of ground truth, sharing and extension of such resources in terms of ground truth, features, algorithms and tools etc. The Workshop will discuss and formulate action plan towards these goals.

This workshop welcomes contributions on the following topics:

- Construction, Unification and Evolution of Corpus
- Framework for sharing of dataset, ground truth, features, algorithms and tools
- Indexing and retrieval for large multimedia collections (including images, video, audio and other multi-modal systems)
- Large-scale video event and temporal analysis over diverse sources
- Automatic machine tagging, semantic annotation and object recognition on massive multimedia collections
- Interfaces for exploring, browsing and visualizing large multimedia collections
- Scalable and distributed machine learning and data mining methods for multimedia data
- Performance evaluation methodologies and standards
- Large-scale copy detection and near-duplicate detection
- Web-scale combined analysis of social and content networks
- Scalable and distributed systems for multimedia content analysis

ACM multimedia 2010 hosts the "Multimedia Grand Challenge" for the 2nd time. Large-Scale multimedia applications are among the potential topics for the Grand Challenge. The availability of a Large-Scale Corpus would effectively boost research in this direction and foster many new applications to the grand challenge for the years to come.

## Paper Submission Guidelines

Submission will follow the main conference guidelines:

A submission must represent new, original research. Papers containing substantially similar material may not be submitted to other venues concurrently with ACM MM10 and the workshop. All the submissions should be prepared in double-column ACM proceedings style. Papers must be submitted as PDF files. Papers exceeding the limits of 6 pages will be rejected without review. To publish the paper in the workshop, one of the authors needs to register and present the paper in the workshop.

Submissions are open: <http://www.edas.info/conference.php?c=9027>

Authors of selected papers will be invited to submit extended versions for publication in a **journal special issue** (under discussion)

## Important Dates

**Extended** Submission of full papers: **July 8<sup>th</sup>**, 2010

Notification of acceptance: **July 20<sup>th</sup>**, 2010

Camera-ready Copy Due: **July 26<sup>th</sup>**, 2010

Workshop: October 29<sup>th</sup>, 2010

## Organizing Co-Chairs

Benoit HUET, EURECOM

Tat-Seng CHUA, National University of Singapore

Alex HAUPTMANN, Carnegie Mellon University