

GUEST EDITOR'S INTRODUCTION

CHENGCUI ZHANG

*The University of Alabama at Birmingham
Birmingham, AL 35294, USA
zhang@cis.uab.edu*

1. Introduction

While applications and delivery forms of multimedia data are virtually limitless, the growing volume of multimedia data such as images, audio and videos on the Internet presents unprecedented challenges in data storage, semantic analysis, indexing, retrieval, annotation, delivery, and management.

In the past decade, there has been a renewed and surging interest on data semantics, which tries to model semantics explicitly rather than relying on symbolic artificial intelligence [1]. For multimedia data, two inherent problems in multimedia semantic understanding are the semantic gap between low-level features and high-level concepts and the subjectivity in the user's perception. The former is due to the inability of computer vision systems to map low-level visual descriptors to high-level conceptual schemas and analogies. The latter is related to the question of how to define the "meaning" or semantics of multimedia data. Unlike traditional types of data such as numbers and texts, the content perceived by the end user can significantly vary from one user to another and is dependent on the act of perception. The perceived content is thus inseparable from the user's mental model of the world, his mood and intention at the time he interprets the multimedia object, and other contexts. That is, the "meaning" of a multimedia object can make sense only in the context of an activity and as part of an activity [1]. To address these issues, methods and approaches have been developed to extract knowledge from multimedia data, including feature space computing, self-organizing maps [2], relationship mining, social multimedia mining [3], etc. Information from various modalities, including images/videos/annotations/geospatial features, etc. can be combined and incorporated into an interactive multimedia system (e.g., a user feedback-enabled system) to enable efficient and effective retrieval and cross mining of multimedia semantics. Rich semantics of multimedia data can be exploited in context-aware computing environment to provide personalized services that are envisioned to be adaptive to the current user context such as location, time and devices, and user preferences.

Despite all the advancements in the field of multimedia semantic analysis and modeling, a few challenges still remain. One of the most pressing issues is efficient indexing techniques for web-scale multimedia collections. There can be many different types of multimedia data one user can generate, like a combination of photos, videos, audio, geospatial locations, which need to be parsed, clustered, indexed, and tagged in order to be browsable and searchable. An interactive interface is required to assist the user in manipulating the data and search tasks efficiently and in the personalization of search needs. Such a system is envisioned to be self-evolving, adaptive to the user's needs by observing the user's data usage pattern (e.g., auto-tagging, object recognition). Another key challenge is scalable machine learning and data mining methods for extraction of knowledge from large-scale multimedia data. Many existing data mining approaches cannot be performed at real-time, hindering the capability of semantic exploitation of multimedia collections. Integrating information from different modalities and using them collectively in data mining is another challenge — only very limited studies can be found in this area which try to formally encode and model semantics from different modalities and attach them to the multimedia collection. Still, semantic integration of heterogeneous information sources remains very challenging. Last but not least is the need for a novel visualization interface that can visualize large multimedia collections.

2. About This Special Issue

The articles in this special issue are a selection of the full papers accepted by the 2011 IEEE International Symposium on Multimedia (ISM, 2011), held at Dana Point, California, USA, on December 5–7, 2011. ISM is an international forum for researchers to exchange information regarding advances in the state of the art and practice of multimedia computing, as well as to identify the emerging research topics and define the future of multimedia computing. The technical program of ISM 2011 consists of invited talks, full and short paper presentations, workshops, technical demos, and panel discussions. Among the 24 papers accepted as full length regular papers, three of them that are among the best presentations in the technical sessions were selected for publication in this special issue. The authors were asked to revise their original ISM 2011 presentation and extend it by at least 30% new content. Each submission was rigorously reviewed by a panel of referees selected from the conference program committee and a selected few external reviewers.

The first article entitled “Re-Ranking Algorithm for Multimedia Retrieval via Utilization of Inclusive and Exclusive Relationships between Semantic Concepts” is authored by Chao Chen, Lin Lin, and Mei-Ling Shyu from University of Miami. Two kinds of relationships between semantic concepts are explored in this paper for re-ranking of multimedia retrieval results, namely the inclusive and exclusive relationships. The inclusive relationship refers to the high co-occurrence relationship between a target concept and a reference concept, while the exclusive relationship represents the low co-occurrence relationship. Multiple reference concepts are used to

capture the different aspects of the target concept that are otherwise not available with one single reference concept. Correlation is exploited between attributes and the co-occurrence class built from reference concepts and used to improve the retrieval accuracy for the target concept.

The article by Fausto C. Fleites, Shu-Ching Chen, and Kasturi Chatterjee, entitled “A Semantic Indexing Structure for Multimedia Retrieval,” addresses a critical issue in multimedia retrieval — how to build an efficient index structure that provides not only efficient access to multimedia databases but also accommodates semantic gap and user perception subjectivity? By incorporating high-level affinity information during the retrieval process into the construction of a tree-based index structure, AH+-tree, a height-balanced and distanced-based tree structure, is proposed and implemented to address this problem. This index structure has been thoroughly tested on range and nearest neighbor queries and demonstrated superior I/O efficiency without a detrimental impact on real-time costs of the retrieval process.

The third article by Shogo Kumagai, Keisuke Doman, Tomokazu Takahashi, Daisuke Deguchi, Ichiro Ide, and Hiroshi Murase, entitled “Speech Shot Detection from Broadcast News Videos,” won the “Best Student Paper” award at ISM 2011 in its original form. In this extended version, they proposed a method for discriminating genuine speech shots from narrated shots (where the speaker is not the subject who appears in the shot). Their approach goes beyond face recognition and localization by moving one step further into the exploration of lip-voice co-occurrence and the intra- and inter-shot features. They compared their algorithm with eight other existing methods and achieved the highest overall accuracy of 0.871.

Acknowledgments

The editor wishes to thank all of the authors, reviewers and publisher. Without their excellent work, and the devotion of their valuable time, this special issue would not have been possible.

References

- [1] S. Santini, Semantic computing in multimedia, in *Proc. 2011 ACM Multimedia*, Scottsdale, AZ, USA, 2011, pp. 631–632.
- [2] H. Zheng, P. Cunningham and A. Tsymbal, Learning multiple linear manifolds with self-organizing networks, *International Journal of Parallel, Emergent and Distributed Systems, Special Issue on Multimedia Data Mining* **22**(6) (2007) 417–426.
- [3] G. Lappas, From web mining to social multimedia mining, in *Proc. 2011 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Kaohsiung, Taiwan, 2011, pp. 336–343.