

Research Statement for Annual Review 2012

- *Jialie Shen*

Research Motivations and Problems

Modern developments in digital media and mobile computing technologies have made transferring and storing large amounts of multi/rich media data (e.g. text, images, music, video and their combination) more feasible and affordable than ever before. Meanwhile, rapid growth of online media content has been changing the way how we share, consume and manage the content about our daily life. Unfortunately, the state of the art information system techniques to process, analyze and manage the rich/multimedia are still in their infancy.

The focus of my research is the study of technologies related to large scale of rich/multimedia data management and knowledge discovery. Particularly, I am currently interested in various techniques of information retrieval and content analysis. I also have been working on applying the media analysis techniques to address management issues in business firms since 2007. The areas stem from a number of research fields – information retrieval, computer vision, data mining, pattern recognition and machine learning, marketing, finance, among others. The fundamental research challenges include 1) generating the effective media descriptor for effectively modeling or annotating; 2) retrieving and analyzing the interest media objects efficiently and accurately; 3) assessing the performance of relative systems in a scalable and comprehensive way; 4) developing intelligent search schemes and algorithms for novel domain applications (e.g., video sequences) and 5) developing and validating the economic aware media analysis scheme. The main goal of my research is to assist effective and efficient access, management and knowledge discovery over large volume of multi/rich media data, which is the complex mixtures of visual, audio and text contents and important experience for daily live of humans and business world.

Current Research

My undergoing research efforts can be categorized into three main streams. They include,

Large Scale Media Information Retrieval and Data Management

Empowered by advances in information technologies, recent decades have witnessed the ever-increasing amounts of media stream data available from many different domain applications. Consequently, there are growing efforts to develop computing technique for media information retrieval and data management. The goal of the research effort is to develop technical advancements to enable general consumers to enjoy rich set of functionalities for the easier and better access media information. At the same time, the project aims to acquire fundamental knowledge about multimedia search engine performance via large scale empirical study. Thus, we develop four sets of fundamental techniques to facilitate effective large scale media retrieval and information management,

- Semantic event and concept detection and annotation [6]
- Indexing structure for fast information retrieval process [2][4]
- Comprehensive and accurate annotation techniques [9][11]

- A set of fast evaluation methods to assess media search engine performance from different aspects

Although the proposed work is fundamental and will contribute to the advancement of knowledge, the results will have immediate practical application in search engine industry at both national and international levels. Recently, I was invited as a panellist for discussing “Web-Scale Media Content Analysis and Retrieval” at the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC'10). During the panel discussion, I also present a few my research results in the related domain.

The study is partially supported by the Lee foundation Fellowship and the MOE Tier 2 grant (No: R263-000-617-112).

Personal Media Information Management under Mobile Environment

The pervasiveness and rapid growth of mobile and wireless devices (e.g., mobile phone and PDAs) hold the great potential to change the way how information about our daily activities could be accessed and consumed. Taking advantages of ubiquitous wireless infrastructure, related mobile applications can create virtual space enabling us to capture, search, memorize, share and recall personal information about daily life in the form of richmedia (e.g., image, video, sound and the associate text) under the dynamic mobile environment (e.g., driving and walking). Unfortunately, a number of challenges still hinder related technical advancement. A fundamental concern is that processing and understanding the data are generally required to consume amounts of computing resource and energy – all of which are scarce on the mobile devices. In this project, the ultimate research goal is to develop a cloud based framework called E-Recall for effective personal information management, searching, sharing and archiving via integrating the needs of mobile end users [11]. The aim of our approach is to allow high quality analysis, exchanging, publishing and searching process over large scale of personal rich media information collections. It can seamlessly combine technical developments such as mobile search, cloud computing and multimodality integration. The framework can be applied to many different domain applications (e.g., auto or semi auto blogging).

The study is supported by Mobile plus Cloud Computing Theme Research Program Award - Microsoft Research. Currently, we are developing a research prototype for demonstrating basic research ideas and major domain applications.

Intelligent Media Analysis for Addressing Management Challenges (with a Focus on Marketing)

In this research, we study the value relevance of top management’s marketing strategy focus via the comprehensive analysis of large scale enterprise documents. It is motivated by the key observation that the influence of marketing in the boardroom has been marginalized. The present study makes the case for a stronger marketing strategy focus of top management. We argue that – contrary to many boards’ current practice – top managers should devote more attention to marketing strategies. Towards this goal, we draw on the attention-based view of the firm from the organizational behaviour literature, the framework of market-based assets, and other concepts from the marketing strategy literature to examine how and when marketing strategy focus of top management leads to better financial performance of firms. To facilitate fast and accurate firm

performance related document processing, we plan to develop a set of document classification schemes for deep text analysis processes.

This research is a cooperative study with Professor Sundar Bharadwaj from Goizueta Business School, Emory University and Professor Stefan Worm from HEC Paris - School of Management, France. It is partially supported by the Merlion Programme.

Future Research Directions

Advanced media retrieval and analysis techniques play very important role in modern information management systems. In the next phase of my research study, I will continue my current efforts in this domain and aim to achieve a series of records with higher real impacts. At the mean time, I believe that the techniques can be applied to address many management challenges and be very useful to unleash the economic value of media data in different domain. Consequently, my future research will focus on three major directions. They are,

Automatic Multimedia Data Annotation

Recent literatures show that text queries are currently more effective and intuitive than content-based queries for searching multimedia data. However, since textual description of multimedia objects is acquired via manual annotation, this process is not only labor intensive but also hard to scale to a large collections. While significant progress has been achieved for developing automatic annotation algorithms, most of the existing systems still adopt a simple classification approach, and apply machine learning classifiers directly on low level media features. Consequently, they suffer the shortcomings of (1) poor accuracy, (2) lack of comprehensive evaluation results and the associated analysis based on large scale datasets, and (3) incomplete content representation, arising from the lack of the multimodal and structural information integration. This project aims to develop a set of novel techniques to annotate multimedia data objects with tags. One of the main novelties is to effectively integrates both multimodal and structure information in the representation of multimedia signal. We also will plan to develop multi-instance learning framework for exploring the correlation among tags to generate a comprehensive annotation.

Intelligent Indexing Structure for Retrieval Scalability Improvement

A massive amount of media data is being generated on daily basis; it is currently in the order of terabytes or petabytes per day, and will continue to increase. To achieve fast access to such a large volume of media data such as image, music, video, etc., efficiency is an important issue and an efficient indexing structure is essential to scale the data space. Various kinds of multidimensional access methods have been proposed in recent years. Most of them are designed based on a principle called "feature transformation", which is emerging as an important search paradigm in multimedia information systems. The key idea is to extract the features (usually in the form of a multidimensional vector) from each multimedia objects (e.g., image or audio) in the database and then to map features into points in a high-dimensional feature space. The distance between two feature points is frequently used as a measure of similarity between multimedia objects. Once the distance or similarity function is defined for the multidimensional feature space, k-nearest neighbor search can be used to retrieve the objects. The retrieval results need to satisfy the criteria specified in a given query.

However, it is extremely difficult to achieve accurate query processing using indexing structure constructed based on low level features only. The main reason behind is that low level features have very limited capability to describe complex semantic meaning and high level similarity notions due to the “semantic gap”. On the other hand, multidimensional feature vectors usually have high dimensions (e.g., some image feature vectors could have up to 100 dimensions) and creating a generalized high-dimensional index that can handle hundreds of dimensions is still a difficult problem without proper solution to date. Motivated by the concerns above, this project aims to develop a novel technique to effectively and efficiently indexing multimedia data. Its architecture is designed based on a "classify-and-indexing" principle and contains two main components 1) a novel semantic-sensitive classification process to identify a target object's category; and 2) an efficient searching process based on local indexing structure for each semantic category. To assess the performance of this framework from different aspects, we also plan to develop a comprehensive evaluation procedure and a set of test collections/queries.

Reference

- [1] **J. Shen** and J. Shepherd. "Efficient Benchmarking of Content-based Image Retrieval via Resampling". Proceedings of the 14th ACM International Conference on Multimedia (ACM MM'06), 2006. (Full research paper)
- [2] B. Cui, L. Liu, C. Pu, **J. Shen**, K-L. Tan. "QueST: Querying Music Databases by Acoustic and Textual Features". Proceedings of the 15th ACM International Conference on Multimedia (ACM MM'07), 2007. (Author list is in alphabetic order) (Full research paper)
- [3] **J. Shen**, J. Shepherd, AHH. Ngu. "Efficient Content Based Music Retrieval with Multiple Acoustic Feature Composition", IEEE Transactions on Multimedia (IEEE TMM), 2006. (Full research paper)
- [4] **J. Shen**, J. Shepherd, B. Cui, K-L. Tan. "A Novel Framework for Efficient Automated Singer Identification in Large Music Databases". ACM Transactions on Information Systems (ACM TOIS), 2009. (Full research paper)
- [5] **J. Shen**. "Stochastic Modeling Western Paintings for Effective Classification". Pattern Recognition, 2009. (Full research paper)
- [6] **J. Shen**, D. Tao, X. Li. "Modality Mixture Projections for Semantic Video Event Detection. IEEE Transactions on Circuits and Systems for Video Technology, (IEEE T-CSVT), 2008. (Full research paper)
- [7] **J. Shen**, M. Wang, S. Yan, H. Pang, and X-S Hua. "Effective Music Tagging through Advanced Statistical Modeling", ACM SIGIR Conference, 2010. (Full research paper)
- [8] **J. Shen**, S. Yan, X-S. Hua. "The E-Recall Environment for Cloud based Mobile Rich Media Data Management", ACM Workshop on Mobile Cloud Media Computing (MCMC'10), 2010. (Full research paper)
- [9] H. Pang, **J. Shen**, R. Krishnan. "Privacy-Preserving, Similarity-Based Text Retrieval", ACM Transactions on Internet Technology (ACM TOIT), 2010. (Full research paper)
- [10] **J. Shen** and Z. Cheng. "Personalized Video Similarity Measure", to appear in ACM Multimedia Systems Journal (ACM MMSJ), 2012. (Full research paper) .
- [11] **J. Shen**, M. Wang, S. Yan, X.-S. Hua. Multimedia Tagging: Past, Present and Future. ACM International Conference on Multimedia (ACM MM'11), 2011 (conference tutorial)