

# NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

## PH.D DEFENCE - PUBLIC SEMINAR

**Title:           Geo-referenced Video Retrieval: Text Annotation and Similarity Search**

Speaker:       Ms Yin Yifang

Date/Time:     6 May 2016, Friday, 10:00 AM to 11:30 AM

Venue:         Executive Classroom, COM2-04-02

Supervisor :   Dr Roger Zimmermann, Associate Professor, School of Computing

### Abstract:

Advanced technologies in consumer electronic products have enabled individual users to record, share and view videos with mobile devices. With the volume of videos increasing tremendously on the Internet, fast and accurate video search and annotation have become urgent tasks and have attracted much research attention. However, video search and management operations are typically supported by either the low-level visual features or the manual textual annotations. Those approaches often suffer from low recall as they are highly susceptible to changes in viewpoint, illumination, and noisy tags. By leveraging geo-metadata, more reliable and precise search results can be obtained. The geographic metadata is one of the important kinds of contextual information. Due to the ubiquity of sensor-equipped smartphones, it has become increasingly feasible for users to capture videos together with the geographic information, for example the location and the orientation of the camera. Such contexts create new opportunities for the organization and retrieval of geo-referenced videos.

This dissertation studies the geographic information use in video annotation and retrieval. Since raw sensor data collected is often noisy, we first preprocess the geo-metadata by building a comprehensive model to reduce the errors in GPS and compass readings. The proposed approach can effectively provide more accurate geo-metadata for downstream applications such as tagging and search. For video annotation, we propose to leverage crowdsourced data from social multimedia applications that host tags of diverse semantics to build a spatial-temporal tag repository. In particular, we retrieve the necessary data from several social multimedia applications, mine both the spatial and temporal features of the tags, and then refine and index them accordingly. Consequently, the tag repository we built acts as the input to our previous auto-annotation approach which we extend in several ways for better integration with the new vocabulary. For video landmark retrieval, we present the Geo Landmark Visibility Determination (GeoLVD) approach which computes the visibility of a landmark based on intersections of a camera's Field-of-View (FOV) and the landmark's geometric information available from geographic information systems and

services. We compare our method with the content-based spatial pyramid matching approach combined with two advanced coding methods: sparse coding and locality-constrained linear coding. By analyzing their strength and weakness, we further integrate the visual and geographic information to achieve improvements. For video similarity search, we propose a novel video description which consists of (a) determining the geographic coverage of a video based on the camera's FOV and a pre-constructed geo-codebook, and (b) fusing video spatial relevance and region-aware visual similarities to achieve a robust video similarity measure. Toward a better encoding of a video's geo-coverage, we also construct a geo-codebook by segmenting a map into a collection of coherent regions. The experimental results show that the proposed techniques achieved significant improvements over its competitors, especially with fine-grained and accuracy-enhanced geographic metadata.