In recent years, the amount of multimedia data such as images, texts, and videos have been growing rapidly on the Internet. Motivated by such trends, this thesis is dedicated to exploiting hashing-based solutions to reveal multimedia data correlations and support intra-media and inter-media similarity search among huge volumes of multimedia data.

We start by investigating a hashing-based solution for audio-visual similarity modeling and apply it to the audio-visual sound source localization problem. We show that synchronized signals in audio and visual modalities demonstrate similar temporal changing patterns in certain feature spaces. We propose to use a permutation-based random hashing technique to capture the temporal order dynamics of audio and visual features by hashing them along the temporal axis into a common Hamming space. In this way, the audio-visual correlation problem is transformed into a similarity search problem in the Hamming space. Our hashing-based audio-visual similarity modeling has shown superior performances in the localization and segmentation of sounding objects in videos.

The success of the permutation-based hashing method motivates us to generalize and formally define the supervised ranking-based hashing problem, and study its application to large-scale image retrieval. Specifically, we propose an effective supervised learning procedure to learn optimized ranking-based hash functions that can be used for large-scale similarity search. Compared with the randomized version, the optimized ranking-based hash codes are much more compact and discriminative. Moreover, it can be easily extended to kernel space to discover more complex ranking structures that cannot be revealed in linear subspaces. Experiments on large image datasets demonstrate the effectiveness of the proposed method.

We further studied the ranking-based hashing method for the cross-media similarity search problem. Specifically, we propose two optimization methods to jointly learn two groups of linear subspaces, one for each media type so that features' ranking orders in different linear subspaces maximally preserve the cross-media similarities. Additionally, we develop this ranking-based hashing method in the cross-media context into a flexible hashing framework with a more general solution. We have demonstrated through extensive experiments on several real-world datasets that the proposed cross-media hashing method can achieve superior cross-media retrieval performances against several state-of-the-art algorithms.

Lastly, to make better use of the supervisory label information, as well as to further improve the efficiency and accuracy of supervised hashing, we propose a novel multimedia discrete hashing framework that optimizes an instance-wise loss objective, as compared to the pairwise losses, using an efficient discrete optimization method. In addition, the proposed method decouples the binary codes learning and hash function learning into two separate stages, thus making the proposed method equally applicable for both single-media and cross-media search. Extensive experiments on both single-media and cross-media retrieval tasks demonstrate the effectiveness of the proposed method.

Major: Computer Science

Educational Career:
Bachelor's of Automation, BS, 2010, Huazhong University of Science and Technology
Master's of Computer Science, MS, 2015, University of Central Florida

Committee in Charge:
Kien Hua, Chair, Computer Science
Guojun Qi, Department of Computer Science, University of Central Florida
Haiyan Hu, Department of Computer Science, University of Central Florida
Morgan Wang, Department of Statistics

Approved for distribution by Kien Hua, Committee Chair, on November 30, 1999.
The public is welcome to attend.