

Abstract

Many image retrieval schemes generally rely only on a single mode, (either low level visual features or embedded text) for searching in multimedia databases. In text based approach, the annotated text is used for indexing and retrieval of images. Though they are very powerful in matching the context of the images but the cost of annotation is very high and the whole process suffers from the subjectivity of descriptors.

In content based approach, the indexing and retrieval of images is based on the visual content of the image such as color, texture, shape, etc. While these methods are robust and effective they are still bottlenecked by semantic gap. That is, there is a significant gap between the high-level concepts (which human perceives) and the low-level features (which are used in describing images). Many approaches (such as semantic analysis) have been proposed to bridge this semantic gap between numerical image features and richness of human semantics

Semantic analysis techniques were first introduced in text retrieval, where a document collection can be viewed as an unsupervised clustering of the constituent words and documents around hidden or latent concepts. Latent Semantic Indexing (LSI), probabilistic Latent Semantic Analysis (pLSA), Latent Dirichlet Analysis (LDA) are the popular techniques in this direction. With the introduction of bag of words (BoW) methods in computer vision, semantic analysis schemes became popular for tasks like scene classification, segmentation and content based image retrieval. This has shown to improve the performance of visual bag of words in image retrieval. Most of these methods rely only on text or image content.

Many popular image collections (eg. those emerging over Internet) have associated tags, often for human consumption. A natural extension is to combine information from multiple modes for enhancing effectiveness in retrieval.

The enhancement in performance of semantic indexing techniques heavily depends on the right choice of number of semantic concepts. However, all of them require complex mathematical computations involving large matrices. This makes it difficult to use it for continuously evolving data, where repeated semantic indexing (after addition of every new image) is prohibitive. In this thesis we introduce and extend, a bipartite graph model (BGM) for image retrieval. BGM is a scalable datastructure that aids semantic indexing in an efficient manner. It can also be incrementally updated. BGM uses

tf-idf values for building a semantic bipartite graph. We also introduce a graph partitioning algorithm that works on the BGM to retrieve semantically relevant images from a database. We demonstrate the properties as well as performance of our semantic indexing scheme through a series of experiments.

Then , we propose two techniques: Multi-modal Latent Semantic Indexing (MMLSI) and Multi-Modal Probabilistic Latent Semantic Analysis (MMpLSA). These methods are obtained by directly extending their traditional single mode counter parts. Both these methods incorporate visual features and tags by generating simultaneous semantic contexts. The experimental results demonstrate an improved accuracy over other single and multi-modal methods.

We also propose, a tri-partite graph based representation of the multi model data for image retrieval tasks. Our representation is ideally suited for dynamically changing or evolving datasets, where repeated semantic indexing is practically impossible. We employ a graph partitioning algorithm for retrieving semantically relevant images from the database of images represented using the tripartite graph. Being "just in time semantic indexing", our method is computationally light and less resource intensive. Experimental results show that the data structure used is scalable. We also show that the performance of our method is comparable with other multi model approaches, with significantly lower computational and resources requirements