Abstract

Recent explosive growth in images and videos accessible to any individual on the Internet have made automatic management the prime choice. Contemporary systems used tags but over time they were found to be inadequate and unreliable. This has brought content based retrieval and management of such data to the fore front of research in the information retrieval community.

Content based retrieval methods generally represent visual information in the images or videos in terms of machine centric numeric features. These allow efficient processing and management of a large volume of data. These features are primitive and thus, are incapable of capturing the way humans perceive visual content. This leads to a semantic gap between the user and the system, resulting in poor user satisfaction. User centric techniques are needed which will help reduce this gap efficiently. Given the ever expanding volume of images and videos, techniques should also be able to retrieve in real time from millions of samples. A practical image retrieval approach in summary is expected to perform well on most of the following parameters, (1) acceptable accuracy, (2) efficiency, (3) minimal and non-cumbersome user input, (4) scalability to large collections (millions) of objects, (5) support interactive retrieval and (6) meaningful presentation of results. Most of the noted efforts in CBIR literature have focused primarily on providing answers to only subsets of the above. A real world system, on the other hand, requires practical solutions to nearly all of them. In this thesis we propose our solutions for an image retrieval application, keeping in mind the above expectations.

In this thesis we present our system for interactive image retrieval from huge databases. The system has a web-based interface which emphasizes ease of use and encourages interaction. It is modeled on the query-by-example paradigm. It uses an efficient B+-tree based dimensional indexing scheme for retrieving similar ones from millions of images in less than a second. Perception of visual similarity is subjective. Therefore, to be able to serve these varying interpretations the index has to be adaptive. Our system supports user interaction through feedback. Our index is designed to support changing similarity metrics using this feedback and is able to respond in sub-second retrieval times. We have also optimized the basic B+-tree based indexing scheme to achieve better performance when learning is available.

Content based access to images requires the visual information in them to be abstracted into some numeric features. These features generally represent low level visual characteristics of the data like colors, textures and shapes. They are inherently incapable of weak and cannot represent human perception of visual content, which has evolved over years of evolution. Relevance feedback from the user has been widely accepted as a means to bridge this semantic gap. In this thesis, we propose an inexpensive, feedback driven, feature relevance learning scheme. We estimate iteratively improving relevance weights for the low level numeric features. These weights capture the relevant visual content and are used to tune the similarity metric and iteratively improve retrieval for the active user. We propose to incrementally memorize this learning across users, for the set of relevant images in each query. This helps the system in incrementally converging to the popular content in the images in the database. We also use this learning in the similarity metric to tune retrieval further. Our learning scheme integrates seamlessly with our index making interactive accurate retrieval possible.

Feature based learning improves accuracy; however it is critically dependent on the low level features used. On the other hand, human perception is independent of it. Based on the underlying assumption, that user opinion on the same image remains similar over time and across users, a content free approach has recently become popular. The idea relies on collaborative filtering of user

interaction logs for predicting the next set of results for the active user. This method performs better by virtue of being independent of primitive features but suffers from the critical cold start problem. In this thesis we propose a Bayesian inference approach for integrating similarity information from these two complimentary sources. It also overcomes the critical shortcomings of the two paradigms. We pose the problem as that of posterior estimation. The logs provide a priori information in terms of co-occurrence of images. Visual similarity provides the evidence of matching. We efficiently archive and update the co-occurrence relationships facilitating sub-second retrieval.

Studies have shown that the user refines his query based on the results he is shown. Studies have also shown that quality of retrieval improves with the effectiveness of learning acquired by the system. Presenting the right images to the user for his feedback can thus result in the best retrieval. A set of results which are similar to the query in different ways can help the user narrow down to his intended query at the earliest. This diversity cannot be achieved by similarity retrieval methods. In this thesis, we propose to efficiently use skyline queries for effectively removing conceptual redundancy from the retrieval set. Such a diversely similar set of results is then presented to the user for his feedback. We use our indexing scheme to extract the skyline efficiently, a computationally prohibitive process otherwise. User's perception changes with the results, so should the nature of the diverse set. We propose the idea of preferential skylines to handle this. We use the user preference, based on feedback, to adapt the retrieval to the user intent. We reduce the diversity and include more similarity from the preferred attributes. Thus, we are slowly able to tune the retrieval to match the user's exact intent. This provides improved accuracy and in far fewer iterations.

We validate all the ideas proposed in the thesis with experiments on real natural images. We also employ synthetic datasets for other computational experiments. We would like to mention that in spite of the considerable improvements in accuracy with our learning approaches, the effectiveness our solutions is still limited by the features used for encoding the human visual perception. Our methods of learning and other optimizations are only a means to reduce this gap. We present extensive experimental results with discussions validating different aspects of our expectations from the proposed ideas, throughout this thesis.