A Scalable Graph-Based Semi-Supervised Ranking System for Content-Based Image Retrieval

Xiaojun Qi, Department of Computer Science, Utah State University, Logan, UT, USA
Ran Chang, Department of Computer Science, Utah State University, Logan, UT, USA

ABSTRACT

The authors propose a scalable graph-based semi-supervised ranking system for image retrieval. This system exploits the synergism between relevance feedback based transductive short-term learning and semantic feature-based long-term learning to improve retrieval performance. Active learning is applied to build a dynamic feedback log to extract semantic features of images. Two-layer manifold graphs are then built in both low-level visual and high-level semantic spaces. One graph is constructed at the first layer using anchor images obtained from the feedback log. Several graphs are constructed at the second layer using images in their respective cluster formed around each anchor image. An asymmetric relevance vector is created for each second layer graph by propagating initial scores from the first layer. These vectors are fused to propagate relevance scores of labeled images to unlabeled images. The authors’ extensive experiments demonstrate the proposed system outperforms four manifold-based and five state-of-the-art long-term-based image retrieval systems.

Keywords: Anchor Images, Content-Based Image Retrieval, Graph-Based Semi-Supervised Ranking System, Long-Term Learning, Semantic Features, Short-Term Learning

1. INTRODUCTION

With the rapidly growing number of digital images found on the Internet and housed in digital libraries, the need for effective and efficient tools to manage large image databases has grown dramatically. Specifically, the development of efficient image retrieval systems to find images of interest in this haystack of data has become an active research area in recent years (Thomee, 2010).

Content-based image retrieval (CBIR) techniques (Datta et al., 2008; Lew et al., 2006) are viable solutions to find desired images from multimedia databases and have evolved significantly since the early 1990s. They make use of low-level visual image features (e.g., color, texture, shape, etc.) instead of keywords to represent images, where each feature can be automatically and consistently extracted without human intervention. Consequently, they overcome the limitations entailed by text-based...
image retrieval, which include the large amount of manual labor required to annotate each image in the database, and the inconsistency among different annotators in perceiving the same image. However, as the ranking of retrievals is calculated based on selected image features, the retrieval accuracy may be unsatisfactory due to the semantic gap between low-level visual features and high-level semantic concepts. This semantic gap exists because objects of the same type do not have the same visual representation. For example, images of similar semantic content may be scattered far away from each other in the feature space, while images of dissimilar semantic content may share similar low-level features. To bridge the semantic gap, a great deal of research work has been focused on developing effective relevance feedback (RF) techniques (Liu et al., 2007; Zhou & Huang, 2003), which utilize users’ interaction to learn better representation of images as well as the query concept. RF, as an interactive search technique, has been used in CBIR systems to repeatedly modify the query descriptive information (feature, matching models, metrics or any meta knowledge) as response to the users’ feedback on retrieved results. Therefore, it learns the query close to its optimal and returns more user-desired images (i.e., improves the retrieval precision) after each round.

Most existing RF techniques use short-term learning or intra-query learning to find out images that are relevant to the user’s query in a retrieval session. Representative short-term learning techniques include query updating (e.g., query reweighting, query shifting, and query expansion) and statistical learning techniques (e.g., inductive learning and transductive learning) (Qi et al., 2011). However, query updating methods (Munesawang & Guan, 2004) do not fully utilize the information embedded in feedback images and therefore cannot achieve satisfactory retrieval results. Inductive learning methods (Tong & Chang, 2001; Wu & Yap, 2006) yield degraded retrieval results when the chosen classifier is trained with insufficient labeled training samples. Moreover, these two categories of techniques ignore the manifold structure of image features. Therefore, the latest trend has been moving towards RF-based transductive learning, which explores the relationship of all database images in the feature space and propagates ranking scores of labeled images to unlabeled images via a weighted graph. To this end, He et al. (2006) propose a generalized manifold-ranking-based image retrieval (gMRBIR) algorithm, which works well for any query image inside or outside the database. Their proposed algorithm represents images and their relationships as a graph and propagates labeled information among images via the graph structure. It further exploits the distribution of unlabeled images to improve the retrieval performance. Wan (2007) proposes to apply the MRBIR algorithm to non-overlapping equal-sized blocks of each database image and fuse the ranking scores of all blocks in the image as the final retrieval score of each image. Cai et al. (2007) incorporate a locality preserving regularizer into the manifold structure to learn a classification function in the image manifold. They then apply the user’s RFs to update the manifold structure for better classification. Bian and Tao (2010) combine the biased discriminative Euclidean embedding with the manifold regularization-based item to discover a more accurate manifold structure for better classification. Geng et al. (2012) propose an ensemble manifold regularization framework to implicitly estimate hyperparameters involved in the regularization to better explore the intrinsic manifold structure of the image database. All above transductive methods achieve better retrieval precision in each iterative step. However, they do not apply users’ accumulated historical RF information to improve the manifold graph. They also cannot run on a computer when the number of images in the database reaches a certain level due to the use of several large square matrices. Furthermore, all these short-term learning techniques cannot capture the semantic meaning of an image and therefore cannot achieve satisfactory retrieval results. They also cannot remember users’ historical feedback and therefore cannot utilize it in future retrievals.
Towards Fusion of Textual and Visual Modalities for Describing Audiovisual Documents

An Improved Security 3D Watermarking Method Using Computational Integral Imaging Cryptosystem