Information Processing and Management 52 (2016) 571–591

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

A semantics and image retrieval system for hierarchical image databases

Shreelekha Pandey[a], Pritee Khanna[a,*], Haruo Yokota[b]

[a] PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Dumna Airport Road, Jabalpur 482005 Madhya Pradesh, India

[b] Graduate School of Information Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

ARTICLE INFO

Article history:
Received 11 September 2015
Revised 2 December 2015
Accepted 10 December 2015
Available online 6 January 2016

Keywords:
Content based image retrieval
Semantic assignment
Clustering
Visual search space
Indexing of visual features

ABSTRACT

This work presents a content based semantics and image retrieval system for semantically categorized hierarchical image databases. Each module is designed with an aim to develop a system that works closer to human perception. Images are mapped to a multi-dimensional feature space, where images belonging a semantic are clustered and indexed to acquire its efficient representation. This helps in handling the existing variability or heterogeneity within this semantic. Adaptive combinations of the obtained depictions are utilized by the branch selection and pruning algorithms to identify some closer semantics and select only a part of the large hierarchical search space for actual search. So obtained search space is finally used to retrieve desired semantics and similar images corresponding to them. The system is evaluated in terms of accuracy of the retrieved semantics and precision-recall curves. Experiments show promising semantics and image retrieval results on hierarchical image databases. The results reported with non-hierarchical but categorized image databases further prove the efficacy of the proposed system.

1. Introduction

During last two decades image retrieval emerged as a promising technology which also received a widespread interest of media. Researchers from various fields such as image processing, pattern recognition, computer vision, psychology, and many more have joined hands together and put their sincere efforts into understanding the real world implications, applications, and constraints in image retrieval. Many techniques were developed as a result of their collaborative work (Datta et al., 2008; Deselaers et al., 2008; Liu et al., 2007; Vassilieva, 2009). Initially, image retrieval systems worked on a text-based framework, then came the content-based image retrieval (CBIR), and now the focus has been shifted to automatic image annotation (AIA). In CBIR systems, image contents are visually interpreted for retrieval and a visual feature space is generated to search similar images. These systems try to emulate human perception through visual distances in the form of similarity measures defined on color, texture, and shape features. The major challenges in this domain are image database for testing, feature extraction and indexing techniques along with distance measures corresponding to human-visual system, query processing, user friendliness, and visual semantic gap (Datta et al., 2008). In such a state of affairs, till now CBIR community does not have universally acceptable algorithms to characterize human vision in the context of object recognition and image retrieval.

* Corresponding author. Tel.: +91 7612794222.
E-mail addresses: shreelekha@iiitdmj.ac.in (S. Pandey), pkhanna@iiitdmj.ac.in, priteekh@gmail.com (P. Khanna), yokota@cs.titech.ac.jp (H. Yokota).

http://dx.doi.org/10.1016/j.ipm.2015.12.005
0306-4573/

It is time to think about more sophisticated and robust models and algorithms to index, retrieve, organize, and interact with image data.

Most of the existing CBIR systems/approaches limit their contributions to one or two of these issues. In an attempt to effectively address these issues, this work proposes a prototype for a full-fledged semantics and image retrieval system working on a large image database. The system is developed especially for semantically categorized hierarchical image databases and is fully automatic irrespective of the size of the database. Semantic based categorization of an image database results in a hierarchical tree structure with categories and subcategories of images at various levels. Visual features of images in such a database form a semantic based hierarchical search space on which the proposed system works. The core of the system consists of an effective way of traversing the hierarchical search space (to reduce the search space/time), clustering performed within categories (to reduce semantic gap), and indexing of clusters (to further reduce the search time). A user friendly interface is developed for experimentation. Experiments are performed with two sets of query images (inside and outside database) on a hierarchical image database (e.g., ImageNet), and also on a flat database (e.g., WANG), to analyze the efficacy of the system for semantic assignment as well as image retrieval.

The manuscript is organized as follows. Section 2 presents a brief review of CBIR systems. Section 3 presents the proposed content based semantics and image retrieval system and discusses its modules including feature extraction, clustering of image data, indexing, and adaptive branch selection on clustered and indexed data. Results are discussed in Section 4 and Section 5. Finally, Section 6 concludes the work.

2. Related work

CBIR systems map images to a multidimensional space formed with color, texture, and shape features (Deselaers et al., 2008). A suitable similarity function is used to measure the visual similarity of two images. The choice of features may depend on the application and the database used for retrieval. Research shows that the systems using a combination of multiple visual features are more accurate in terms of precision and recall (Hiremath & Pujari, 2007; Lin et al., 2009; Wang et al., 2011).

Blobworld (Carson et al., 2002), CIRES (Iqbal & Aggarwal, 2002), MARS (Rui et al., 1997), MIT's Photobook (Pentland et al., 1994), NeTra (Ma & Manjunath, 1999), QBIC (Niblack et al., 1993), SIMPLIcity (Wang et al., 2001), VisualSEEk (Smith & Chang, 1997), WebSEEk (Chang et al., 1997) are some well-known CBIR systems. Some unnamed systems are also available in literature (ElAlami, 2011; Lin et al., 2009; Shrivastava & Tyagi, 2015; Youssef, 2012). These systems focus on image retrieval solely to improve the efficiency of CBIR system. The most popular among these, e.g., IBM's QBIC (Query By Image Content) applies all three features and uses R*-trees for enhanced speed. It accepts queries based on example images, user-constructed sketches or/and selected color, and texture patterns (Niblack et al., 1993). Blobworld (Carson et al., 2002) first segments images to find regions that have an associated color and texture descriptors indexed using a tree. At the time of querying, instead of entire image, description of a few regions is used to facilitate feasible large-scale retrieval. Similarly, NeTra identifies homogeneous regions and computes their color, texture, shape, and spatial location information. It provides the flexibility to specify properties of the region of interest, and the system searches for images with the similar regions in the database (Ma & Manjunath, 1999). VisualSEEk is unique as it provides joint color/spatial querying capability. It uses color set back-projection technique to extract color regions. The colors, sizes, spatial locations, and relationships of regions are used to compare images (Smith & Chang, 1997). SIMPLIcity (Semantics sensitive Integrated Matching for Picture LIbraries) also represents an image as a set of regions, which are characterized by color, texture, shape, and location (Wang et al., 2001). The system first classifies an image into semantic category and then applies semantically-adaptive feature extraction and searching methods. This categorization narrows down the searching range and helps in enhanced retrieval. CIRES (Content-based Image REtrieval System) combines image structure with color and texture features without putting much emphasis on the segmentation or detailed object descriptions (Iqbal & Aggarwal, 2002). Another visual information system, WebSEEk is designed for Web-based environments. It collects, analyzes, indexes, and searches visual information on the Web (Chang et al., 1997). These systems do not put much emphasis on relevance feedback to enhance the performance. However, MARS (Multimedia Analysis and Retrieval System) uses an integrated relevance feedback architecture to address semantic gap and human perception. Feedback mechanism is implemented at feature representation as well as similarity measure levels to improve efficiency (Rui et al., 1997). Photobook (Pentland et al., 1994) works on any commonly used color, texture, and shape features. It uses an interactive learning agent to select and combine features for a given task, and also provides a library of matching algorithms.

Many of these systems do not make use of clustering (ElAlami, 2011; Lin et al., 2009; Shrivastava & Tyagi, 2015; Youssef, 2012), but a few like Blobworld, MARS, and QBIC used clustering to determine regions within an image, i.e., segmentation. Some CBIR systems, like SemQuery (Sheikholeslami et al., 2002), RIME (Chang et al., 1999), CLUE (Chen et al., 2005) have used clustering to group images prior to image retrieval. Researchers have also utilized clustering for the purpose of image annotation and indexing (Datta et al., 2008; Ober et al., 2007), and object discovery (Tuytelaars et al., 2010). Clustering is also applied on videos to automatically learn actions and then use them for action categorization and localization (Jiang et al., 2013; Niebles et al., 2008).

In order to retrieve images resembling to human perception, SemQuery (Sheikholeslami et al., 2002), RIME (Chang et al., 1999), and CLUE (Chen et al., 2005) used the concept of clustering successfully. SemQuery is a semantics-based clustering and indexing approach that uses a set of heterogeneous features which are merged by means of a hierarchical clustering.