Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 2554–2570
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Region-based image retrieval with high-level semantics using decision tree learning

Ying Liu, Dengsheng Zhang*, Guojun Lu

Gippsland School of Information Technology, Monash University, Vic. 3842, Australia

Received 30 March 2007; received in revised form 1 October 2007; accepted 10 December 2007

Abstract

Semantic-based image retrieval has attracted great interest in recent years. This paper proposes a region-based image retrieval system with high-level semantic learning. The key features of the system are: (1) it supports both query by keyword and query by region of interest. The system segments an image into different regions and extracts low-level features of each region. From these features, high-level concepts are obtained using a proposed decision tree-based learning algorithm named DT-ST. During retrieval, a set of images whose semantic concept matches the query is returned. Experiments on a standard real-world image database confirm that the proposed system significantly improves the retrieval performance, compared with a conventional content-based image retrieval system. (2) The proposed decision tree induction method DT-ST for image semantic learning is different from other decision tree induction algorithms in that it makes use of the semantic templates to discretize continuous-valued region features and avoids the difficult image feature discretization problem. Furthermore, it introduces a hybrid tree simplification method to handle the noise and tree fragmentation problems, thereby improving the classification performance of the tree. Experimental results indicate that DT-ST outperforms two well-established decision tree induction algorithms ID3 and C4.5 in image semantic learning.

Keywords: RBIR; Semantic image retrieval; Decision tree learning; Semantic template; CBIR

1. Introduction

Conventional content-based image retrieval (CBIR) systems index images by their own visual contents such as color, texture and shape. A number of commercial products and experimental prototype systems have been developed, such as QBIC [1], Photobook [2], Virage [3], VisualSEEK [4], Netra [5] and SIMPLIcity [6]. Since the representation of images at region level is closer to human perception system [7], many of these systems support region-based image retrieval (RBIR) [1,2,4–6]. However, due to the 'semantic gap' between low-level image features and the richness of user semantics, the performance of CBIR is still far from user's expectations [8]. Therefore, to further improve retrieval accuracy, a CBIR system should reduce the 'semantic gap' between low-level image features and human semantics [8,9]. Another advantage of semantic-based image retrieval is that it supports query by keywords or textual descriptions which is more convenient for users.

The state-of-the-art techniques for reducing the 'semantic gap' can be roughly classified into five categories [10]. (1) Using machine learning tools to associate low-level image features with high-level semantics [7,11–14]. For example, in Ref. [13], Fei-Fei et al. developed an incremental Bayesian algorithm to learn generative models of object categories and tested it on images of 101 widely diverse categories. (2) Introducing relevance feedback (RF) into retrieval loop for continuous learning of users' intention [7,15–19]. Considering the interaction with the details in an image (such as points, lines and regions), Nguyen and Worring proposed a framework to dynamically update the user- and context-dependent definition of saliency based on RF [15]. (3) Exploring domain knowledge to define ontology for image annotation [16,20]. For instance, Guus et al. designed an annotation strategy to search photograph collections using the background knowledge contained in

* Corresponding author. Tel.: +61 3 99026772.
E-mail addresses: dengsheng.zhang@infotech.monash.edu.au (D. Zhang), guojun.lu@infotech.monash.edu.au (G. Lu).

0031-3203/$30.00 © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.12.003

ontology [20]. (4) Making use of multiple information sources such as the textual information obtained from the Web and the visual content of images for Web image retrieval [21–23]. (5) Generating semantic templates (STs) to support semantic-based image retrieval [22,24,25]. Chang et al. introduced the idea of semantic visual template (SVT) to link low-level image feature to high-level concepts for video retrieval [25]. Many systems exploit one or more of the above techniques to implement high-level semantic-based image retrieval.

Among the above-mentioned techniques, machine learning tools such as support vector machine (SVM), artificial neural networks (ANN) and Bayesian networks (BN) are often used for image semantic learning [7,11–14,17–19]. In Ref. [14], Cao and Fei-Fei propose to use a Bayesian learning model built upon regions obtained through the latent topic allocation. However, the clarity of the model is less than SVM. It is difficult to judge the reliability of the system since the model is complicated by a number of conditional probabilities involved, most of which are not defined in the paper. Results are only given based on small data sets of single object images. In addition, the results depend on how well the latent topics are defined, which have to be manually labelled or collected. In image retrieval, SVM is often used in semantic learning when there are feedbacks from the user [17–19]. In our system, however, higher-level concepts are learned directly from low-level features, no additional information or user feedback is assumed at the present. SVM is good for two class classification, for multiple classes, there needs multiple machines which is difficult for integration. Another difficulty of SVM is to determine an appropriate kernel which is crucial for nonlinear image data.

Many researchers have found that decision tree (DT) learning such as ID3, C4.5 and CART perform well in data classification [26–32]. DT learning is an extensively researched area for classification tasks and has great potential in image semantic learning. One of the beauties of DT is its semantic interpretation power which is a natural simulation of human learning. None of the aforementioned learning tools can derive human comprehensible rules as DT does. The other major advantage of DT is that it is intuitive with highly hierarchical clarity. Compared with other learning methods, DT learning is also simpler and robust to incomplete and noise data due to post-pruning techniques [26,27]. These advantages put DT well competitive with those machine learning techniques in literature.

The difficulty in applying DT induction to image semantic learning lies in the difficulty of discretizing image features [33]. To benefit image learning process, discrete image feature values should correspond to meaningful conceptual names. Algorithms like ID3 [26] require the value of the input attributes to be discrete. Some algorithms have been designed to handle continuous-valued attributes [28,29,34]. For example, C4.5 [28] uses a minimal entropy heuristic to find binary-cuts for each attribute in order to discretize continuous attributes. Based on the understanding that multi-interval discretization is better than binary discretization and that it can lead to more compact DTs, entropy-based multi-interval discretization method is introduced to find multi-level cuts for each attribute [29,34]. However, these generally designed algorithms usually do not provide physical meaning in image feature space. It has been reported that although C4.5 can handle continuous attributes, it does not work as well in domains with continuous attribute values as in domains with discrete attribute values [35].

Another challenge in DT induction is to keep the tree simple by controlling its size and complexity, since a cumbersomely large tree leads to misclassifications [30]. A tree may expand haphazardly due to *noise* (i.e., feature mislabelling and class mislabelling), *fragmentation* (i.e., leaf nodes that represent very few instances), a missing attribute or a missing class label. The most commonly used approaches to reducing tree size include pre-pruning and post-pruning. Pre-pruning techniques are computationally efficient as they prevent the expansion of tree by imposing a stopping criterion that halts the tree induction process, before the complete training data set has been used for tree induction [26]. However, the challenge in pre-pruning is the selection of the stopping criterion that might terminate the induction of tree prematurely, resulting in inconsistent classification performance [31]. This has lead to a wider interest in post-pruning algorithms which simplify the tree after it has been induced using the complete training set. Post-pruning processes the induced tree and removes/replaces unnecessary branches and unknown outcomes. While various existing tree simplification approaches [30] attempt to address tree growth problems, to our knowledge there does not exist any pruning technique that suits well for image semantic learning.

The limitations discussed above provide the motivation to develop an RBIR system with high-level semantics derived using DT learning. Every image in the database is segmented into different regions, represented by their color and texture features. Spatial location and shape feature are not used, as they are not well-defined for regions in natural scenery images, which is the focal semantic domain in this paper. To associate low-level region features with high-level image concepts, we propose a DT-based image semantic learning algorithm, named DT-ST. The DT induction process is based on the concept of top-down induction of DTs. DT-ST is different from other DT induction algorithms in that it introduces a simple and effective image feature discretization method and employs a hybrid tree simplification methodology. For image feature discretization, a set of STs is generated for the concepts defined in our database. A ST is the representative feature of a concept and is calculated from the low-level features of a collection of sample regions. DT-ST converts low-level color/texture features into color/texture labels, thus avoiding the difficult image feature discretization problem. For tree simplification, DT-ST employs a hybrid of pre-pruning and post-pruning techniques in order to resolve the noise and tree fragmentation problems. As a result, the tree grows in a well controlled manner and the classification performance is improved. Tested on a few data sets of natural scenery image regions, the results indicate that DT-ST outperforms other tree induction algorithms such as ID3 and C4.5 by providing higher classification accuracy.

Based on the decision rules derived by DT-ST, each region in an image is associated with a high-level concept. Our system supports both query by specified region and query by keyword. For query by region, we assume that every image contains a