J. Vis. Commun. Image R. 30 (2015) 136–152

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Short Communication

An effective use of adaptive combination of visual features to retrieve image semantics from a hierarchical image database ☆

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

[a] PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Dumna Airport Road, Jabalpur 482-005, M.P., India
[b] Graduate School of Information Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

A R T I C L E I N F O

Article history:
Received 7 November 2014
Accepted 20 March 2015
Available online 31 March 2015

Keywords:
Image semantics
Semantic similarity
Visual similarity
Hierarchical image database
Semantic based categorization
Visual features
Visual search space
Search space pruning

A B S T R A C T

Correlating semantic and visual similarity of an image is a challenging task. Unlimited possibilities of objects classification in real world are challenges for learning based techniques. Semantics based categorization of images gives a semantically categorized hierarchical image database. This work utilizes the strength of such database and proposes a system for automatic semantics assignment to images using an adaptive combination of multiple visual features. 'Branch Selection Algorithm' selects only a few subtrees to search from this image database. Pruning Algorithms further reduce this search space. Correlation of semantic and visual similarities is also explored to understand overlapping of semantics in visual space. The efficacy of the proposed algorithms analyzed on hierarchical and non-hierarchical databases shows that the system is capable of assigning accurate general and specific semantics to images automatically.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

A person can easily infer some semantics from an image. For example, it is easy for us to infer semantics like *Sledding*, *Sports*, *Person*, *Grandparent and grandchildren* and many more from the image given in Fig. 1(a). We are adept to correlate visual similarity to semantic similarity and have natural instinct to group 'similar objects' in categories, and 'similar categories' to 'super-categories' [1]. As a result, for the image in Fig. 1(b), one can easily infer the semantic *Tiger*, followed by *Carnivore* and *Animal*. However, it is difficult for a computer to infer such semantics from an image file. A computer easily computes low level features based on color, texture, and shape. Fig. 2 shows some images with their low level color features and high level semantics. Content Based Image Retrieval (CBIR) systems try to emulate human vision through visual similarity obtained in terms of low level image features to interpret images [2,3]. The lack of coincidence between low-level visual data and high level semantics of images is known as semantic gap [4]. Development of universally acceptable algorithms to reduce semantic gap and characterize human vision for object recognition and image retrieval are in progress [5].

2. Problem statement

Empowering computers to distinguish object categories in visual as well as in semantic space, is a challenging task. Obtaining knowledge of specific semantics is not straightforward even for humans many times. Consider the sunflower images of seven categories, namely *Swamp*, *Common*, *Giant*, *Showy*, *Maximilian*, *Prairie*, and *Jerusalem* in Fig. 3. A semantically categorized hierarchical image database may help us to automatically derive these semantics. The semantic based categorization of images leads to a hierarchical tree structure having images of different categories at various levels. This categorization may help us to understand the correlation between visual features and semantic of categories (e.g. *Animal*, *Vegetable*, *Fruit*, etc.), which may further be utilized to provide specific semantics of the image. In an attempt to correlate visual and semantic similarities, this work aims to derive as exact semantics as possible at a moderate search cost by exploring only some branches of image tree.

3. Related work

Learning algorithms for limited number of concepts are extensively used on flat image databases to reduce semantic gap [6,7]. In a statistical modeling approach for automatic linguistic indexing of pictures, each of the 600 concepts is represented by a two-dimensional multi-resolution hidden Markov model and is

☆ This paper has been recommended for acceptance by M.T. Sun.

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

http://dx.doi.org/10.1016/j.jvcir.2015.03.010
1047-3203/© 2015 Elsevier Inc. All rights reserved.

Fig. 1. Images with semantics as inferred by human.

				
Color	Upper	Blue	Blue	Green
	Middle	Grey/White	Grey/White	Green/Blue/Black
	Lower	Green	Green	Green
Semantics		Structure, Housing, Living Quarters	Sports, Golf, Outdoor Game	

Fig. 2. Some images with low level color features and high level semantics.

trained with categorized images [6]. Generative probabilistic models for 101 object categories are learned through Bayesian incremental algorithm using a few training images for quick learning [7]. Visual recognition from semantic segmentation of photographs is learned to achieve 70.5% region-based recognition accuracy on a 21-class database [8]. Gaussian Mixture Models learned from bags of localized features of images with common semantic label are pooled into a density estimate for the corresponding semantic class [9]. Using the class densities, a minimum probability of error rule is used for image annotation and retrieval. For the same purpose, ExpectationMaximization algorithm and asymmetric PLSA (Probabilistic Latent Semantic Analysis) learning based on textual/visual information of images are also used to learn a model [10].

Optimization and estimation techniques are used in an automatic real-time image annotation system to represent objects by bags of weighted vectors grouped on D2-clustering [11]. Hypothetical Local Mapping is utilized to develop a generalized mixture modeling technique for non-vector data. In another real-time image annotation approach key phrases of similar images are mined for candidate annotations [12]. The approach is scalable and robust to outliers as it does not require any training data. Image metadata and parametric dimensions were used to obtain a set of rules in a decision tree based automatic semantic annotation approach [13]. The system is developed with 3231 manually labeled images and tested on 1,00,000+ Web images outside the training database.

Label correlations are explored to develop a two-dimensional active learner for image classification, and an adaptation algorithm is used to update the model [14]. Another label transfer based non-parametric system used a SIFT flow algorithm to retrieve dense scene correspondences from a fully annotated large database [15]. These correspondences are used to integrate multiple cues to recognize query images. A generic multiview latent space Markov network developed to relate image features and abstract concepts maximizes the likelihood of multiview data and minimizes a prediction loss on the labels from side information [16]. An optimal Image-tag relation matrix consistent to the observed tags and the visual similarity is obtained through a semi-supervised algorithm [17]. In another approach, an automatic news image caption generation system learned extractive and abstractive surface realization models from weakly labeled data in an unsupervised fashion [18].

Unlike traditional hypergraph learning, weights of hyperedges are adaptively learned in many works to improve the performance [19–22]. The size of neighborhood is varied to generate a set of hyperedges, where weights are optimized by means of a regularizer [19]. Click data is integrated with the system to reduce the semantic gap [19,21]. The images are also classified by combining information from labeled views [22]. A probability distribution constructed using high-order relationship is estimated through hypergraph. Also, visual and textual information are utilized for social image search [23]. The weights of hyperedges are learned to enhance the effects of informative visual words and tags. Learning employs a set of pseudo-relevant samples based on tags. Recently, The generative approach to identify visual neighborhood in training image set, is refined adaptively by discriminative hyperplane tree classifier [24].

A few more techniques exist in literature but unrestricted concepts in the real world limit the power of learning based approaches in general. Hierarchical structures are also used for the purpose of image retrieval [25], object recognition [26,27], indexing [28] or codebook generation [29]. A tree structure, built by identifying various objects in the set of training images and arranging them on different levels depending on the relationship among the identified objects, is explored to get the desired results. Instead of focusing on learning techniques, Khanna et al. used a hierarchical image database to assign efficient semantics to a given image [30]. With the aim of deriving image semantics, a large hierarchical image database is used to establish a correlation between visual and semantic similarities. The present work extends the concept and aims to derive semantics with high precision in the reduced search cost.

The rest of the paper is organized as follows. Section 4 gives an insight of related image databases. Proposed methodology is explained in Section 5. Section 6 discusses the representation of semantic categories of images in visual space. 'Branch Selection Algorithm' given in Section 7, is followed by pruning approaches in Section 8. Section 9 summarizes results and discussion on related issues. Finally, Section 10 concludes the work along with future directions.

4. Image databases

The nature of image database influences the design and performance of semantics retrieval algorithms. For decades, researchers used self-collected images to show their results. Later, many domain specific databases having thousands of uncategorized images, e.g., WANG, UW, IRMA 10000, ZuBuD, and UCID came into

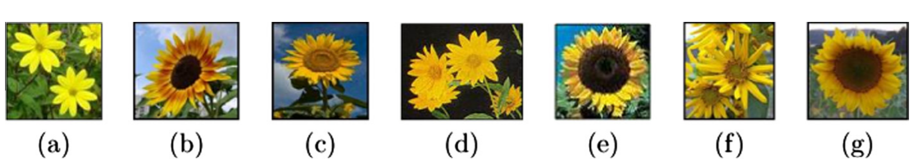

Fig. 3. Images from seven categories of *Sunflower* as per ImageNet. (a) Swamp. (b) Common. (c) Giant. (d) Showy. (e) Maximilian. (f) Prairie. (g) Jerusalem.

Download English Version:

https://daneshyari.com/en/article/529150

Download Persian Version:

https://daneshyari.com/article/529150

Daneshyari.com