Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 70 (2007) 1040–1050
NEUROCOMPUTING
www.elsevier.com/locate/neucom

A new image classification technique using tree-structured regional features

Tommy W.S. Chow*, M.K.M. Rahman

Dept of Electronic Engineering, City University of Hong Kong, Hong Kong

Received 12 November 2004; received in revised form 26 January 2006; accepted 26 January 2006
Communicated by T. Heskes
Available online 10 October 2006

Abstract

Image classification is a challenging problem of computer vision. Conventional image classification methods use flat image features with fixed dimensions, which are extracted from a whole image. Such features are computationally effective but are crude representation of the image content. This paper proposes a new image classification approach through a tree-structured feature set. In this approach, the image content is organized in a two-level tree, where the root node at the top level represents the whole image and the child nodes at the bottom level represent the homogeneous regions of the image. The tree-structured representation combines both the global and the local features through the root and the child nodes. The tree-structured feature data are then processed by a two-level self-organizing map (SOM), which consists of an unsupervised SOM for processing image regions and a supervising concurrent SOM (CSOM) classifier for the overall classification of images. The proposed method incorporates both global image features and local region-based features to improve the performance of image classification. Experimental results show that this approach performs better than conventional approaches.

Keywords: Feature integration; Region-based image comparison; Image similarity measure; Self-organizing map; Image classification

1. Introduction

Recently, content-based image classification and retrieval received increasing attention through numerous applications [8,10–13,21] in the field of education, entertainment, military, and biomedicine. With the immense growth of computational power and the continuously declining hardware cost, image retrieval, image classification, and pattern recognition have become more demanding in the area of computer vision. The success of solving such problem lies in the issues of object-based image understanding, proper representation of image contents, and suitable learning algorithms. Traditionally, flat image features (fixed length features extracted from whole image), such as color, texture, and their combinations [8,10,12,13,17], are used for image retrieval and classification. Color histogram, which can be extracted efficiently, is the most popular feature. Color histogram is generally insensitive to small changes in camera positions, but it is a very coarse representation in a sense that it discards other important features like textures, shapes, sizes, and positions of the objects. Thus, color histogram lacks spatial information and is sensitive to intensity variation, color distortion, and cropping. As a result, images with similar histogram may exhibit substantially different semantics [14]. To overcome the limitation of color histogram, color layout approach was introduced to partition an image into blocks. Subsequently, the average colors [16], or Daubechies' wavelet coefficients, [25] of those blocks were used as features. The shape, the position, and the texture information can be maintained at a proper resolution, but the retrieval performance is sensitive to the procedures of shifting, cropping, or scaling, because local properties are emphasized. Gabor filter [10] is widely adopted to extract texture features from images and it has been shown to be an efficient approach for performing image retrieval and classification. Combination of color

*Corresponding author. Tel.: +852 27887756; fax: +852 27887791.
E-mail address: eetchow@cityu.edu.hk (T.W.S. Chow).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.01.033

and texture features [17] is shown to be more effective in describing image content compared with color or texture alone. All these flat vector-type image features are excellent in comparing the overall image similarity, but they are still unable to describe the high level and the object-based property of the image content.

A meaningful image representation is the key for performing classification of images. Region-based approaches [3,9,20,21,24] were introduced for a better understanding of image contents. In the region-based approaches, instead of extracting features from the whole image, the image is first decomposed into regions for the subsequent features extraction. The similarity between two images is expressed as all the possible combinations of the similarities among their regions. However, region-based approach has been limited to image retrieval application for two main reasons. First, the comparison of two images is not as straightforward as comparing two feature vectors. Second, traditional neural learning is not directly applicable to such non-flat structured data. Therefore, one query image must be compared with every other image of the database. Tree-structured representation is shown to be an effective approach for image processing and image analysis [19]. There are evidences indicating that region-based image representation can be better encoded using a tree representation. Some successful works can be found in [1,2,15,18,26]. In [15,26], binary space partition (BSP) tree representation is used for coding images in terms of segmented regions. Recent works in [1,2,18] indicate that the BSP tree-based region-oriented image representation can be effective for image processing, classification, and retrieval. Though BSP provides efficient processing of data, partitioning a region into two sub-regions in BSP does not always provide a meaningful representation of the objects in a real image. Height of the BSP tree grows linearly with the number of objects in the image that lacks clarity in representing image contents.

In this paper, an improved image classification approach is proposed. Image contents are represented by integrating both the global image features and the local region-based features. Fig. 1 briefly illustrates the proposed approach. A two-level tree hierarchically contains all image features, where the root node represents the whole image and the child nodes represent the image regions. Thus, the global and local image features are represented by the root and child nodes, respectively. Two-level self-organizing map (SOM) networks are used to process the tree-structured data. First, all image regions of the whole database are processed by an unsupervised SOM. After the completion of training, image regions are compressed by the positions of winner neurons on the SOM map. The position vectors together with the global image features are then used to classify the images through a concurrent self-organizing maps (CSOM) classifier [11]. Experimental results indicate that the proposed method delivers better results in the application of image classification. It is also worth noting that the proposed method is able to maintain the computational cost at a reasonably low level.

This paper is organized as follows. In Section 2, the representation and the feature extraction of image contents are described. Section 3 elaborates a two-layer SOM-based image classification system. Section 4 presents the experimental results and discussions. Finally, conclusion is drawn in Section 5.

2. Representation of image contents

Feature extraction and representation is important in the sense that they serve as dimensionality reduction for enabling the analysis of image contents. However, features should also reflect the high-level semantic (perceived by human) in addition to visual similarity obtained by global features like color histogram. Thus, different region-based approaches were proposed to provide a better understanding of image semantic. In this work, a composite region-based image representation is developed to integrate both the visual and regional properties of the image content. To extract regional features, JSEG, a color image segmentation method [6], is used. In brief, JSEG first quantizes colors of an image to several representative classes. It then labels pixels with the color classes to form a class map of the image. At last, the image is segmented using multi-scale J-images [6]. Experimental results showed

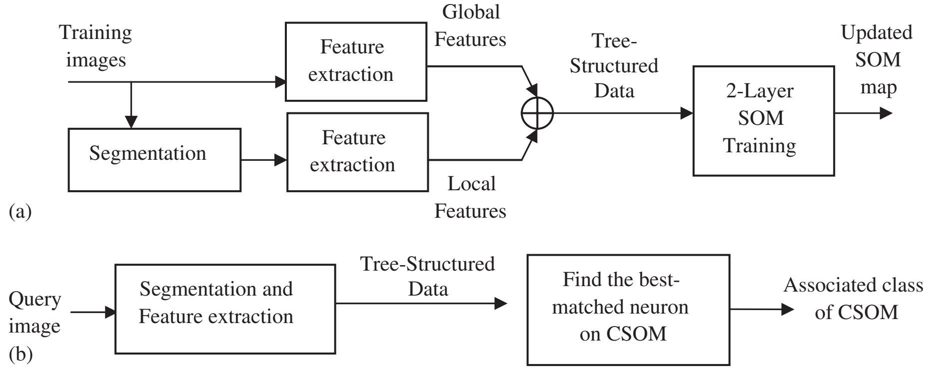

Fig. 1. Overview of the 2-layer SOM-based classification system (a) training phase and (b) classification phase.