Content based image retrieval using tree-structured self-organizing maps

José Alfredo F. Costa and Raquel E. Patiño-Escarcina

Adaptive Systems Laboratory, Center for Technology
Federal University of Rio Grande do Norte, Natal-RN, Brasil
alfredo@dee.ufrn.br, raquel.patino@gmail.com

Abstract

Content-based image retrieval systems are designed to provide effective access to image databases, based on their visual contents and according to a given criteria. This paper focuses the image searching based on descriptors automatically extracted from the images. It is presented a scheme that decomposes the image collection in a hierarchy of clusters using tree-structured self-organizing maps. The present approach uses different sets of features in each level of the hierarchy. Subsets of images are clustered in each level of the tree of maps, associated to regions of neurons, and are used to train subsequent maps in the hierarchy. The tree-structured self-organizing map algorithm may be seen as a recursive image data partition clustering method. Using of this architecture is motivated not only by the reduction of complexity but also as a weak resemblance of the visual information system that process different types of information in different layers of neurons.

1 Introduction

Content-based image retrieval (CBIR) addresses the problem of finding images relevant to the users' information needs from image databases, based principally on low-level visual features for which automatic extraction methods are available [1]. The areas of applications are wide, including e-commerce solutions, space and military applications, medicine, biological image processing and much more.

There are two frameworks for image retrieval (IR): text and content based. Text based IR was introduced in the 1970's and requires manual annotation of images with text descriptors, with high subjectivity of human perception and writing. Content based IR, introduced in the 1990's, explores the use of image features, such as color, texture and shapes, to automatically (or semi-automatically) perform indexation based on the visual content of the images.

This paper uses the strategy of representing images as a set of p features. Each image could be regarded as a point in a p-dimensional feature space. The automatic organization and retrieval of a collection of images can then be posted as a problem of data clustering, which objectives to find a valid organization of the data based on the inherent structure and relationships among the patterns [2]. This paper presents a scheme that decomposes the image collection in a hierarchy of clusters using a structured neural network.

2 Tree-structured self-organizing maps

Self-organizing maps (SOM) has been applied to a variety of tasks including data visualization and clustering [5]. It implements an orderly mapping from a high-dimensional data distribution onto a regular low-dimensional grid of neurons. Once the point density of the neurons approximates the density of data, it is possible to miner clustering information from neuron's relations after its learning.

Tree-structured self-organizing map (TS-SOM) was first described in the 1990's [6] and was made of several SOMs arranged to a tree (fixed) structure. All nodes in the tree (except leaf nodes) have four children. Training is performed one level at time, from top down, with the weights of previous level being used to initialize the next level. A main objective was reducing training time of large maps.

Costa and Netto [3, 4] introduced a different hierarchical structure of self-organizing maps using maps as nodes instead of single neurons. From top to down, maps are automatically segmented by using the *U-matrix* information, which presents relations between neighboring neurons. The partitioning algorithm was based on mathematical morphology segmentation. Clusters of neurons are automatically identified and labeled and generate new sub-maps. Data are partitioned accordingly the label of its associated neurons in each level of the tree. The algorithm may be seen as a recursive partition clustering with multiple prototypes cluster representation, and enables the discoveries of clusters in a variety of geometrical shapes. The height of the tree and the number of maps for a given level are automatically detected.

3 CBIR using TS-SOM and a descriptors hierarchy

The current proposal is to perform clustering in each level of the tree of maps. The difference of this new implementation of TS-SOM to references [3, 4] is the usage of different descriptors sets in levels of the hierarchy (see figure 1). It is motivated by the reduction of complexity (decrease of input data dimensionality) as well as a weak resemblance of the visual information system that, in some parts, process different types of information in different layers of neurons. The training as well specific rules for tree generation is similar to references [3, 4].

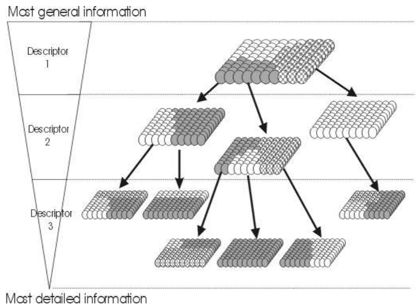

Figure 1: TS-SOM and a descriptors hierarchy

4 Implementation

MPEG-7 [7] is a standard for description of features of multimedia content and includes eleven low level descriptors for images. Results present in this paper uses two sets of descriptors: Color layout (CLD) and edge histogram descriptors (EHD). The CLD is obtained by applying DCT transformation on the 2-D array of local representative colors in YCbCr color space, where each channel is represented by eight bits and averaging each of the three channels separately for 8 image blocks. The EHD represents local edge distribution in an image by dividing the image into 4 sub-images and generating a histogram from the edges present in each of these sub-images. A histogram with 80 bins is obtained, leading thus to a feature vector of 80 dimensions.

The first level on the TS-SL-SOM neural network was trained with feature vectors of the most general descriptors for all images in the dataset. After, this map was segmented in *k* regions; each map in the next level is trained with the feature vector of next descriptor set. This process continues until it reaches the last level in the hierarchy. The two descriptors set could be arranged in two different hierarchies. After several experiments, CLD gave the most general information of images and was used in the first level of the hierarchy. Experiments used to take this decision were made with the same images used for testing the retrieval efficiency.

5 Performance Evaluation

Consider a database D with *N* images obtained from different classes of objects. In most of cases there are no distinct and objective boundaries between different image classes.

Precision (***Pr***) and recall (***Nr***) are used in CBIR system to measure retrieval performance [8]. ***Pr*** is defined as the ratio of the number of relevant images retrieval (***Nr***) to the number of total retrieved images *k*, or ***Pr*** = ***Nr*** / *k*.

6 Experiments and Results

The image database used was the Columbia Object Image Library [9] that contains 100 different classes, each object rotated in 72 different degrees, resulting in 7200 images in the database. For training the neural network, 50 examples of each class were used and the remaining was used for testing. The experiments used map sizes with 10×10 neurons. Figure 2 present the root map after segmentation, showing pictures with descriptors most similar to the weight of neurons. The map was segmented in four regions, leading to a new 4 maps in the second level.

Figure 2: Segmentation of the root map

This approach was compared with other two approaches, *Parallel Descriptors* [1] and *Composed Descriptors* (use of the full vector of features to image retrieval). Tests were conducted with 2200 images and the mean of the precision rate for the first 20 image classes is showed in the figure 3. Figure 4 show that the precision of our approach is better than other approaches. In mean, after 2200 tests, our approach retrieved at least 22 images from the same class of the input image, while the other approaches only retrieves, in mean, 16 and 14 images respectively.

7 Conclusions

A different approach of tree-structured self-organizing map used [3, 4] was presented with the use of a descriptors hierarchy enables different views in different levels of the structure. Comparisons were performed with the composed descriptors and parallel descriptors approaches. Arranging descriptors in a hierarchy gave better overall results. Current research includes the adoption of relevance feedback and rules to automatically define the best order of sets of descriptors appearing in the hierarchy.

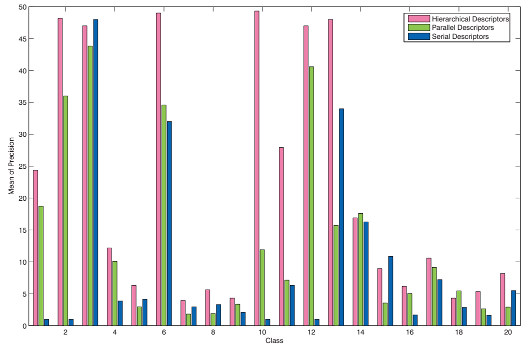

Fig 3. Mean of precision of retrieval images for the first 20 image classes

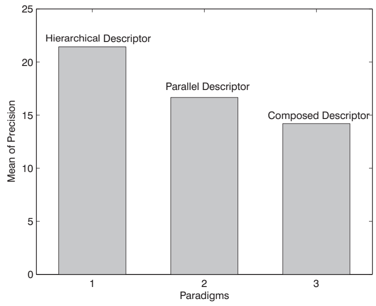

Figure 4. Mean of precision for all classes

References

[1] Koskela, M., Laaksonen, J. and Oja, E. (2004). Use of Image Subset Features in Image Retrieval with Self-Organizing Maps. *Lecture Notes in Computer Science*, vol. 3115, pp. 508–516.

[2] Jain, A.K. et al. (1999). Data clustering: a review. *ACM Computing Surveys*, vol. 31, pp. 264 – 323.

[3] Costa, J.A.F. and Netto, M.A. (2001). A new tree-structured self-organizing map for data analysis. In: *Proc. of the Intl. Joint Conf. on Neural Networks*, pp. 1931– 1936, Washington DC.

[4] Costa, J.A.F. and Netto, M.L.A. (2001). Clustering of complex shaped data sets via Kohonen maps and mathematical morphology". In: *Proc. of the SPIE,* vol. 4384, pp. 16-27, 2001.

[5] Kohonen, T., *Self-Organizing Maps*, 3nd edition, Berlin: Springer-Verlag, 2001.

[6] Koikkalainen, P. and Oja, E. (1990). Self-organizing hierarchical feature maps. In: *Proc. Int. Joint Conf. on Neural Networks, Washington*, DC, v. 11, pages 279-285, Piscataway, NJ.

[7] Salembier, P., and Smith, J. R. (2002). Mpeg-7 multimedia, description schemes. *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, pp. 748–759.

[8] Ma, W.-Y. et al. (1999). Netra: A toolbox for navigating large image databases. *Multimedia Systems*, v. 7,184–198.

[9] Nene, S. A. et al., (1996). Columbia object image library (coil-100). TR CUCS-006-96, URL http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php. Access in Feb., 1996.