Improving Content-based Image Retrieval by Combining Growing Hierarchical Self Organizing Map Classifiers for Color, Shape and Texture Features

Saulo de Tarso O. Rodrigues*, Tiago Oliveira de Albuquerque and Herman M. Gomes
Universidade Federal de Campina Grande, Departamento de Sistemas e Computação
Av. Aprígio Veloso, 882, 58109-970 Campina Grande, PB
saulo@dsc.ufcg.edu.br, tiagooa@lcc.ufcg.edu.br, hmg@dsc.ufcg.edu.br

Abstract

This paper presents the development of a content-based image retrieval system that combines a set of GHSOM (Growing Hierarchical Self Organizing Map)[4] classifiers taking as input color, shape and texture features. Experiments demonstrated combination scenarios that produced better accuracy in image classification.

1. Introduction

Automated search in large image databases is still an open research problem [6]. Most applications found in the literature still present low precision and recall, typically below the levels of 50% and 70% respectively. Moreover, some of the works that obtained better results either considered only a few number of very diverging groups on a large image set or were evaluated using a small image set (e.g. 100 different images) [1, 5]. Such limitations were the starting point for this research, which aimed at studying feature extraction techniques and a classification scheme based on the combination of individual classifiers.

2. Content-based Image Retrieval (CBIR)

A typical CBIR system contains three modules: Feature Extraction; Indexing; and Retrieval. There are several ways to extract image features; this research considered the most common features, which are color, shape and texture.

On current CBIR systems the color characteristic is frequently represented as histograms. For color images, the histograms are commonly calculated for each color component. The color spaces considered in this work were RGB, HSV and YCbCr, each color space provides a different set of features for the same image. Another feature extraction technique involves the description of shapes. In some areas, such as pattern recognition, shapes are important features to identify and distinguish objects [7]. Other than color and texture, the shape is extracted after the image has been segmented in regions or objects. A very used technique in the area of CBIR, which is related to the description of shapes, are Hu's moment invariants [2]. There are several different methods of extraction and representation of textures. LBP (Local Binary Pattern) [3] is a very good example which is invariant to rotation. Another well referenced method for texture representation are the Wavelets, which are capable of representing textures in multiple resolutions and scales.

The classification method adopted in this work was based on GHSOMs, which is a data structure similar to a B-Tree, but it is not usually balanced, nor its nodes have the same size. The nodes of a GHSOM point towards several other similar nodes and hold the tree property. To train each GHSOM, the same image base was used, which had all the color, shape and texture features extracted for each image in the set. Once GHSOM training ends, a tree-like map is created, where images that are similar, relative to a specific features, are represented in nearby neurons. During the recall phase, the specific features extracted from the query image are fed into the corresponding GHSOM. There will be a propagation of the input through the GHSOM up to its leaf nodes, where the most similar group of images to the input image are stored, relative to a given feature.

3. Proposed Approach

Classifier combination is a set of techniques that is proving useful in many application scenarios. Generally, the main advantage of using a combination method is because, since the classification methods are able to overcome the deficiencies of one another, it is possible to improve the overall system's accuracy. The classifier combination method adopted in this work is based on a voting process, where for every image returned by each selected classifier there will be a value associated to it, this value will be used to sort the final result, so the most voted image will be up front

*Saulo de Tarso O. Rodrigues is supported by CAPES

in the result, and considered to be the most similar image to the searched one. To calculate the vote for each image, the quantization error [4] produced by the GHSOM to the image is considered. This value is between zero and one, the closest to zero the more similar the image is. The layer of the GHSOM where the image appears is also used in the calculation. The following equation is used to calculate the vote for each returned image:

$$v_i = \sum_{g}^{G} (1 - QE_{g,i}) \cdot (\frac{LC_{g,i}}{LT_g}) \qquad (1)$$

where i is the image, G is the count of used GHSOM's, $QE_{g,i}$ indicates the quantization error of the image for a given GHSOM, $LC_{g,i}$ indicates the layer of the image for a given GHSOM and LT_g is the total number of layers of a given GHSOM, which results on a double precision value related to every retrieved image.

4. Experimental Results and Analysis

In order to test the proposed approach and classifier combination, part of Microsoft Research Cambridge Object Recognition Image database, version 1.0, has been used. This is a labeled database with 800 images, divided in 16 sets of 50 images each. Six GHSOM networks that were chosen to evaluate classifier combination were the ones which presented the best individual results. The tests first considered all possible combinations from C$(6, 1)$ (isolated results) to C$(6, 6)$, which gives a total of 63 tests. The crude combination does not exclude any results, so the bad images from all the combined GHSOMs will be returned as a valid result. To overcome this deficiency of the combination, a result filtering strategy has been adopted, which consisted of gradually varying the number of images recovered using the classifier combination and recalculating the precision for each number of returned images tested. This showed that, as the allowed number of returned images is reduced, more bad images start being rejected and better images start being accepted.

In comparison to the simple classification methods, classifier combination resulted on a considerable improvement for precision (see Table 1). The simple classification method that obtained the best precision was the one using the color feature on HSV space and 32 levels per component. Initially the simple classification has a better precision, when considering more than 70 images, but with a reduction on the resulting images, there is a natural reduction in the images that are not considered to be similar and the combination results are greatly superior to the ones of the simple classification. The best combination used HSV 16, RGB 32, LBP and Wavelets, increasing the mean precision in over 17% when the number of returned images is set to five. Even with the simpler combinations, the results are superior when the amount of returned images is under 40. The classification combination showed superior results to the simple classification, where only one feature of the image is used. Table 1 also shows that even the best individual result is not superior to the worst combined result.

Table 1. Experimental Results

Classifier	Precision (%)
WAV	12.50
LBP	18.38
Hu	10.78
RGB 32	17.05
HSV 16	20.33
YCbCr 32	19,55
RGB 32 - LBP	29.38
HSV 16 - RGB 32 - LBP	33.13
HSV 16 - RGB 32 - LBP - WAV	**33.52**
HSV 16 - RGB 32 - YCbCr 32 - LBP - WAV	31.48
HSV 16 - RGB 32 - YCbCr 32 - Hu - LBP - WAV	30.08

5. Conclusion

The main contribution of this research was the development of a combination technique for image classifiers based on GHSOM neural networks, improving the overall precision of a content based image retrieval system. The presented experimental results confirm that using a classifiers combination technique yields a considerable improvement of precision and recall for image retrieval.

References

[1] P. B.G., G. S.K., and B. K.K. Color and shape index for region-based image retrieval. *International Workshop on Visual Form*, 2001.

[2] T. S. Huang and Y. Rui. Image retrieval: Past, present, and future. *Symposium on Multimedia Information Processing*, 1997.

[3] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[4] A. Rauber, D. Merkl, and M. Dittenbach. The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. *Neural Networks, IEEE Transactions on*, 13(6):1331–1341, 2002.

[5] E. Stian. Cbir distance measures and genetic programming. *Undergraduate Theses from Norwegian Classification of Images using Color*, 2006.

[6] R. Torres. Recuperao de imagens por contedo. *Revista Computao Brasil - Sociedade Brasileira de Computao*, 2006.

[7] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.