Image Classification Based on Hierarchical Temporal Memory and Color Features

[1, 2]R. Škoviera, [1]I. Bajla

1 Institute of Measurement Science, Department of Theoretical Methods, Slovak Academy of Sciences, Bratislava, Slovakia
2 Faculty of Mathematics, Physics, and Informatics, Department of Applied Informatics, Comenius University, Bratislava, Slovakia
Email: rado.skoviera@gmail.com

Abstract. *The research in the domain of content-based image retrieval (CBIR) is concentrated on several problems, among which two problems are essential: – development of efficient image classification algorithms, and selection of suitable image features. Recently, a biologically inspired Hierarchical Temporal Memory (HTM) network demonstrated promising results in image classification tasks. The focus of this paper is to explore possibilities of this network to be applied to CBIR. In particular, we study the performance of the HTM network when, instead of conventional grey-level images and features, color features are used. The results of our experiments show that using color texture features, defined for a reduced range of color quantization values, performs comparably well as the grey-scale image features.*

*Keywords**:** Hierarchical Temporal Memory (HTM); Content-Based Image Retrieval, Image Color Features*

1. Introduction

The rise of digital techniques for image acquisition and ubiquity of internet connections in recent decade evoked the creation of a huge number of image databases. Searching/retrieving images from such databases, which are suitable for particular purposes, become a frequent and important task in many application areas, e.g. the so-called content-based image retrieval (CBIR) techniques and systems [1]. However, though a number of methods have been proposed and explored in recent years, no satisfying general solutions still exist. The core problems are: i) definition of suitable image features for image content representation, and ii) efficient clusterization of images represented by specific features. Independently, a progress in development of novel models for computational neuroscience has been achieved in the recent decade. One of such models is a biologically inspired Hierarchical Temporal Memory (HTM) network, proposed by George and Hawkins [2], which manifested encouraging results in the field of grey-level image classification problems [3, 4]. In this paper we present a novel application of the HTM network to the problem of CBIR task, in which the HTM has been extended to be able to process color features instead of only grey-level ones.

2. Subject and Methods

Image Features

In CBIR tasks a textual description is usually absent, instead, various image features (descriptors) are used to find images similar to the given query image. Different features may account for particular properties of images from a database. They are grouped into four basic types: (a) local features, (b) shape features, (c) color features, and (d) texture features. Recently, the attention of researchers has been focused on color features and a combination of color and texture features [5]. Color of pixels can be represented in different color spaces, however, for the given image database, no apriori best-suited color space for image discrimination is known.

Selection of Color Features (Color Spaces)

Before comparing the performance of color features we first tested the HTM network with grey-scale features. More specifically, we used simple grey-scale images, and two texture features based on grey-level pixels – Grey-Level Co-occurrence (GLC) and Grey-Level Difference (GLD) [6]. For color features, we selected simple images converted to one of the selected color spaces and then fed to HTM a vector composing of individual color components for each pixel. As an alternative to the two grey-scale texture features, we used Color Co-occurrence Matrix (CCM), in particular, the Reduced Space variant of the CCM defined in [5]. The standard CCM is defined for a couple of color components C_i and C_j (i, j <1, 3> for three-component spaces). It is computed as $Q \times Q$ matrix, where Q is the number of levels used to quantify the value of color. After specification of the displacement vector $\boldsymbol{d}$ = [dx, dy], the value of C_i is determined as the i-th component of the color of a pixel at the given position [x, y]. The value of C_j is then determined as the j-th component of the color of a pixel at the position [$x + dx$, $y + dy$]. The cell of the resulting matrix with the coordinates [C_i, C_j] is incremented by a constant value. For the RSCCM reduced number of quantization levels Q is used. Similarly to [5], in our experiments we used 3×3 isotropic pixel neighborhood instead of a single displacement vector, and the number of quantization levels Q=8. This means that the co-occurrence for every pixel is calculated with all neighboring. We selected six possible color component combinations C_1&C_2, C_1&C_3, C_2&C_3, C_1&C_1, C_2&C_2, C_3&C_3. Then, for each image patch to be fed as an input to the HTM network, we calculated six RSCCMs for the above color combinations. Based on the results of preliminary testing, we concentrated ourselves on four color spaces: *RGB, I1I2I3*, *UVW*, and *Lab* color spaces [6].

Hierarchical Temporal Memory

An HTM network can consist of several levels of elementary nodes which implement the same learning and inference algorithms. HTM has a hierarchical tree-like structure; its zero (sensory) level receives input vectors (vectorized image features) to the first level of nodes. The top level comprises one node serving as a classifier. The HTM nodes operate in two distinct stages – learning and inference. In the first step of the learning process, the node memorizes the representative spatial patterns from its receptive field, generating thereby a codebook of image patterns. The nodes in the level that is learning can be trained separately or use shared representations. After reaching the requested number of quantization points, the memorization process is finished. The ultimate goal of the HTM learning is to detect correct invariant representations of the input world based on the temporal relations contained in the learning sequence. To achieve this, one needs a frequency of transition events, i.e., co-occurrences of the memorized coincidences in adjacent time instances. A sequence of the input patterns generates a sequence of the n coincidences within the node. The last step of the learning process within each HTM node is the generation of a set of temporal groups. A node that has completed its learning phase can be switched into the inference mode. In this mode, the node produces an output vector for every input pattern provided. This vector indicates the degree of membership of the input pattern into each of the temporal groups. There are two phases of the inference process inference in the "spatial pooler" followed by inference in the "temporal pooler". Finally, we obtain an overall belief vector $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]$ that represents closeness of the input pattern to all memorized coincidences [2]. Such a vector is inferred in the spatial pooler for every input pattern. In the second phase of the inference, the temporal pooler makes use of the learned temporal groups and calculates the output vector for the nodes that are above in the HTM hierarchy. Each individual component of the output vector represents a belief that the vector $\boldsymbol{y}$ comes from a particular temporal group.

HTM for CBIR

The HTM network can be used for transformation of input image feature vectors to a vector space better representing original data, thereby better suited to image classification. This is achieved by increased invariance to affine transformations of the transformed feature vectors. As a result, improved classification accuracy is obtained.

3. Experiments

The proposed features have been tested on images selected from the Caltech-256 database. Ten image categories have been selected, each with 100 images for training and 100 for testing. A random classifier with normal distribution would therefore score a classification accuracy of 10% on this database. All the images have been resized to 128 x 128 pixels due to the network's requirement to have standardized input image sizes. In Fig. 1 examples of color images from the database are shown.

Fig. 1. Examples of images used in the experiment. Each image represents a different image category.

We used our implementation of the HTM network, setup as 1-layer network with node-sharing and node size of 8 x 8 pixels. Thus, the network comprised of 16 x 16 nodes. The codebook was created from the training images using online *k*-means clustering. The process started with just one cluster center (image pattern). A new cluster center was added each time, when none of the existing cluster centers was similar enough to the currently processed pattern. Similarity was calculated using Gaussian and Euclidean distance. The minimum required similarity was 0.95. This procedure has run repeatedly until the required amount of cluster centers was achieved (512 in our case). The temporal learning was implemented as a random walk within the training images, and the temporal groups were created from the time adjacency matrix ([2]) using the agglomerative hierarchical clustering. The classification in the top node was carried out using the *k*NN classifier with $k = 1$. The single layer HTM network can be expected not to perform as well as multi-layer network but the relative difference in performance using different features should be the same. However, the computational complexity of single layer network is much smaller and is therefore more feasible to use it for testing of several image features.

The RSCCM was calculated as described in [5]. The two grey-level texture features, GLC and GLD, require a displacement vector to be defined. This was set to [1, 0]. We tested several variations of the displacement vector, but the results showed no significant variation in the classification accuracy. In the GLC, as well as in RSCCM, we used eight quantization levels.

4. Results

In *Table 1*, the results of our experiment are listed. The lowest row of the table shows classification accuracy values achieved by the 1-NN classifier, incorporated as a top node of the HTM network applied to the different image features specified in the first two rows. The image features are divided into three groups: color patches (the specific color space is mentioned in the second row), the reduced scale color co-occurrence matrix (the specific color space is in the second row), and image features based on grey scale images (image patches, GLC, and GLD). For comparison, we applied the 1-NN classifier to the original image space (i.e., without HTM). The classification accuracy obtained for the same dataset was 39,2% for RGB images and 41,2% for grey scale images. The results show that the best performance was achieved with the GLD feature.

Table 1. Classification accuracy (%) for individual image features.

feature	Color Patches				RSCCM				Grey Scale		
	RGB	I1I2I3	UVW	Lab	RGB	I1IJ2I3	UVW	Lab	Patches	GLC	GLD
CA (%)	42,9	43,1	43,5	45,4	39,2	41,6	43,4	44,4	39,8	40,6	45,7

5. Discussion

We can conclude that the use of the HTM network yields higher values of CA than in the case of classification applied to the original data space (both for grey-level and color images). We have showed that using color features as an input for the HTM is encouraging and it performs comparably well as the features based on grey-scale images. Not all possible values of the parameters of co-occurrence matrices, defined for color features, have been tested. Therefore, our future research will be focused on further examination of possibilities how to improve their performance, e.g., using the RSCCM with higher values of Q and development of a color feature similar to the best performing grey-scale GLD.

Acknowledgement

This work has been supported by the Slovak Grant Agency for Science (VEGA project No. 2/0043/13).

References

[1] Deselaers T, Keysers D, Ney H. Features for Image Retrieval: An Experimental Comparison. *Journal of Information Retrieval* 11, 77–107, 2008. DOI 10.1007/s10791-007-9039-3

[2] George D, Hawkins J. Towards a mathematical theory of cortical micro-circuits. *PLoS Computational Biology* 5(10), 2009. DOI 10.1371/journal.pcbi.1000532.

[3] Štolc S, Bajla I. On the optimum architecture of the biologically inspired Hierarchical Temporal Memory model applied to the hand–written digit recognition. *Measurement Science Revue* 10(2), 28–49, 2010. DOI 10.2478/v10048-010-0008-4.

[4] Štolc S, et al. Pair-wise temporal pooling method for rapid training of the HTM networks used in computer vision applications. *Computing and Informatics,* 31, 901–919, 2012.

[5] Porebski A, Vandenbroucke N, Macaire L. Supervised texture classifi-cation: color space or texture feature selection? *Pattern Analysis and Applications* 16, 1–18. DOI 10.1007/s10044-012-0291-9, 2013.

[6] Lew M S. Principles of visual information retrieval. Springer, 2001.