Vectorization of Content-based Image Retrieval Process Using Neural Network

Hanen Karamti, Mohamed Tmar and Faiez Gargouri
MIRACL, Université de Sfax, Route de Tunis Km 10, B.P. 242, 3021 Sfax, Tunisia

Keywords: Content Based Image Retrieval, Image, Retrieval Model, Neural Network, Query, Vector Space Model.

Abstract: The rapid development of digitization and data storage techniques resulted in images' volume increase. In order to cope with this increasing amount of informations, it is necessary to develop tools to accelerate and facilitate the access to information and to ensure the relevance of information available to users. These tools must minimize the problems related to the image indexing used to represent content query information. The present paper is at the heart of this issue. Indeed, we put forward the creation of a new retrieval model based on a neural network which transforms any image retrieval process into a vector space model. The results obtained by this model are illustrated through some experiments.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

The objective of a Content-based Image Retrieval (CBIR) system is to enable the user to find the targeted images among a collection of images. The general idea behind CBIR is to extract from low-level image features (color, texture, shape) (Schettini et al., 2009), expressed in a numerical way the semantic meaning associated with the image. These features are compared in order to determine similarity between images (Tollari, 2006). This similarity is used to rank a set of images according to an image query.

Primarily designed for improving search quality, the retrieval system should not be limited to simple analysis of the collection and direct matching between queries and images. More elaborate techniques are introduced to analyze and represent images' content, such as the generalized vector space model (Karamti, 2013)(Karamti et al., 2012) and the neural networks (Srinivasa et al., 2006). In CBIR, the vector space model is used in a general way when query and images are represented with feature vectors. The neural network is used in CBIR to organize images in classes (Zhu et al., 2007), it is used to associate images in each class according to probabilities with which they are assigned.

Thanks to these techniques, we propose a new model of content-based image retrieval allowing to integrate theories of neural network on a vector space model, so that the images and queries' matching would be as appropriate and effective as possible.

This paper is organized as follows: we review initially certain CBIR systems in section 2. Then, section 3 describes our suggested search model. Section 4 illustrates the qualitative results, obtained with this model. Finally, section 5 contains conclusion and further research directions.

2 RELATED WORKS

For the last few years, many systems of CBIR have been proposed. The majority of these systems made it possible to navigate within the images database and to display their information requirements through an image query process. These systems relied only on low-level features, also known as descriptors. Several systems belong to this category, for instance the QBIC system (Query By Image Content) of IBM (Flickner, 1997). The Frip system (Finding Areas in the Pictures) (ByoungChul Ko, 2005) proposes to carry out search by areas of interest designed by the users (Caron et al., 2005). The RETIN system (Research and Interactive Tracking) (Fournier et al., 2001) developed at the Cergy-Pontoise university, selects randomly a set of pixels in each image in order to extract their color values. The texture is applied with Gabor filters method (Rivero-Moreno and Bres, 2003). These values are gathered and classified via a neural network. The comparison between images is done through similarity calculation between their features

Karamti H., Tmar M. and Gargouri F..
Vectorization of Content-based Image Retrieval Process Using Neural Network.
DOI: 10.5220/0004972004350439
In *Proceedings of the 16th International Conference on Enterprise Information Systems* (ICEIS-2014), pages 435-439
ISBN: 978-989-758-028-4
Copyright © 2014 SCITEPRESS (Science and Technology Publications, Lda.)

(Tollari, 2006).

Some studies were put forward to change their search spaces, such as color space variation descriptor (Braquelaire and Brun, 1997). Certain research works carried out the minimization of the search scope by calculating the closest neighbors designed to bring together the similar data in classes (Berrani et al., 2002). Thus, image retrieval is carried out by looking for a certain class.

The recent works on CBIR are based on the idea where an image is represented using a bag-of-visual-words (BoW), and images are ranked using term frequency inverse document frequency (tf-idf) computed efficiently via an inverted index (Philbin et al., 2007). Other work is based on the visual features for retrieval such as SIFT (Anh et al., 2010), RootSIFT (Arandjelović and Zisserman, 2012) and GIST (Douze et al., 2009). Others have learnt better visual descriptors (than SIFT) (Winder et al., 2009) or better metrics for descriptor comparison and quantization (Philbin et al., 2010).

The disadvantage of these systems is that the user does not always have an image meeting his actual need, which makes the use of such systems difficult. One of the solutions to this problem is the vectorization technique, which allows to find the relevant images with a query which are missed by an initial search. This process requires the selection of a set of images, known as reference. These references are selected randomly (Claveau et al., 2010), or the first results of an initial search (Karamti et al., 2012) or the centroids of the classes gathered by the K-means method (Karamti, 2013).

All these retrieval systems are based on a query expressed by a set of low-level features. The extracted content influences indirectly the search result, as it is not an actual presentation of the image content.

In order to avoid such problem, we propose in this paper a new retrieval model, which receives in the entry a query designed by a score vector, obtained through the application of an algorithm based on a neural network.

3 PROPOSED APPROACH

In this paper, we present a new CBIR method to transform a content-based image retrieval process to a simple vector space model, which builds the connection between the query image and the result score directly via neural network architecture.

Let $(I_1, I_2 ... I_n)$ a set of images , where each I_i is expressed on the set of low-levels features by $(C_{i1}, C_{i2} ... C_{im})$. A query image Q is therefore a vector designed by $(C_{q1}, C_{q2} ... C_{qm})$ features values. Where C_{qi} is the value of feature i in the query and m is the number of features.

$$I_i = \begin{pmatrix} C_{i1} \\ C_{i2} \\ \vdots \\ C_{im} \end{pmatrix} \qquad Q_q = \begin{pmatrix} C_{q1} \\ C_{q2} \\ \vdots \\ C_{qm} \end{pmatrix}$$

The image retrieval process provides a score vector $S_q = (S_{q1}, S_{q2} \ldots S_{qn})$, where n is the number of images in the collection.

We assume that each image retrieval algorithm is associated with a vector space model providing the same image ranking. This vector space model is characterized by a W matrix $(m \times n)$ where for each query Q associeted with a score vector S, we have:

$$Q * W = S \tag{1}$$

In this paper, we attempt to predict matrix W over a set of queries and their associate score vectors S.

To solve equation 1, we build a neural network containing 2 layers L_Q and L_S (see figure 1). L_Q (resp L_S) is relative to the query features values (resp image scores).

$L_Q = \{n_{q1}, n_{q2} \ldots n_{qi} \ldots n_{qm}\}$ is the input layer of our network, where each neuron n_{qi} represents a feature C_{qi}.

$L_S = \{n_{S1}, n_{S2} \ldots n_{Sj} \ldots n_{Sn}\}$ is the output layer, where each neuron n_{Sj} represents a score S_{qj}.

Figure 1 shows the architecture of our neural network. The application of a learning approach in this neural network on a set of queries led to matrix W, where w_{ij} is given by equation 2:

$$W[i,j] = w_{ij} \qquad \forall (i,j) \in \{1,2...m\} \times \{1,2...n\} \tag{2}$$

$$W = \begin{pmatrix} w_{11} & w_{12} & w_{1j} & \ldots & w_{1m} \\ w_{21} & w_{22} & w_{2j} & \ldots & w_{2m} \\ & & \vdots & & \\ w_{m1} & w_{m2} & w_{mj} & \ldots & w_{mn} \end{pmatrix}$$

w_{ij} values are initialized with random values. These values are then calibrated by using a learning process.

For each query $Q_i = (C_{qi1}, C_{qi2} ... C_{qim})$, we propagate C_{qij} values through the neural network in order to compute S_{qj} scores. Since these scores do not correspond to the expected scores (which are provided by the image retrieval process), we use an error back propagation algorithm to calibrate the w_{ij} weights.

This process is done by algorithm 1, where:

s_j: actual score,

S_j: expected score,

α: learning parameter coefficient,

δ: error rate.

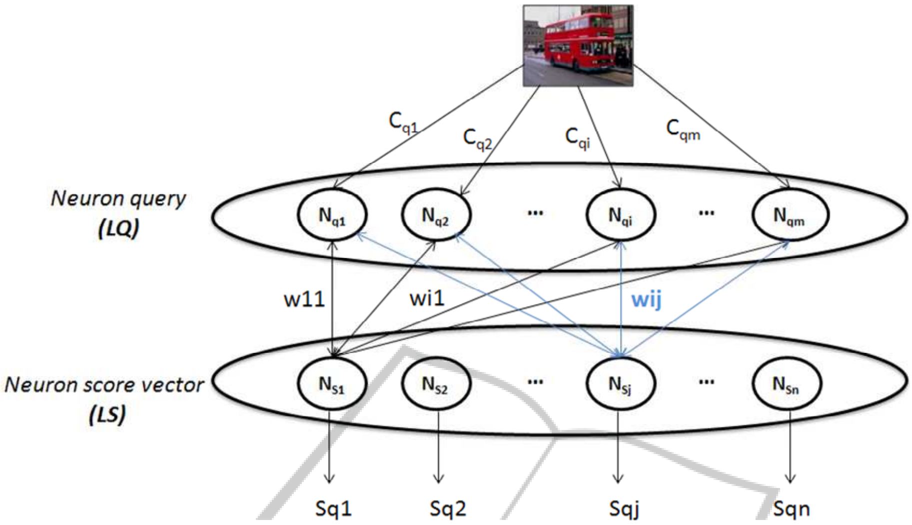

Figure 1: Our neuron network architecture.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

Algorithm 1: The w_{ij} calculating.

$\forall(i,j) \in \{1,2...m\}\{1,2...n\}$
$w_{ij} = 1$
for $eachQ_q = (C_{q1}, C_{q2}...C_{qm})$ **do**
  for $eachI_i = (C_{i1}, C_{i2}...C_{im})$ **do**
    $s_j = \sum_{j=1}^{n} w_{ij} C_{ij}$
    $\delta_j = s_j - S_j$
    $w_{ij} \leftarrow w_{ij} + \alpha\delta_j C_{ij}$
  end for
end for

Once the matrix W is constructed, each query q can be calculated directly without applying an Euclidean metric but by applying of equation 1. Thus, the ranks (scores) of resulting images are obtained as a result of multiplication of query feature vector by a weight matrix W.

4 EVALUATION AND RESULTS

To assess the results of our contribution, we have used a subset of Caltech [1] 205 database. It is constructed by 4000 images. An example of images database is given by figure 2.

To extract the low-level features from images and queries, we used a color descriptor named *CLD* [2] and a texture descriptor called *EHD* [3].

Figure 2: Example of images database used in experiment.

4.1 Results

The performances are measured by the MAP (Mean Average Precision).

The purpose of our retrieval model is to provide the closest results provided by the initial retrieval model. We carried out three tests during the learning phase for construction of W matrix.

At the beginning, we divided our base into two equal parts: training queries (2000 images) and validation queries (2000 images). Then, we change the number of training queries to 3000. The third test comprises 1000 training queries.

Each test is a result of the W matrix. We proceed to integrate each matrix in our retrieval model and we examined the system behaviours for the 10 queries.

Table 1 displays that with Test3, we achieved very close results to those obtained with our initial search system. We also notice the improvement of results obtained after a search with queries (3, 4, 5, 6, 7, 8, 9

[1] http://www.vision.caltech.edu

[2] A color layout descriptor (CLD) is designed to capture the spatial distribution of color in an image. The feature extraction process consists of two parts; grid based representative color selection and discrete cosine transform with quantization.

[3] Edge Histogram Descriptor (EHD) is proposed for MPEG-7 expresses only the local edge distribution in the image.

Table 1: Performance achieved by the new retrieval model with three versions of training dataset.

Queries/MAP	Initial retrieval model	Test 1	Test 2	Test 3
	0.33	0.20	0.14	**0.32**
	0.20	0.20	0.12	**0.19**
	0.10	0.12	0	**0.19**
	0.01	0	0	**0.03**
	0.02	0.02	0	**0.05**
	0.09	0	0	**0.15**
	0.29	0.17	0.12	**0.31**
	0.54	0.35	0.35	**0.56**
	0.09	0.1	0.07	**0.11**
	0.19	0.11	0.08	**0.22**

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

and 10) in comparison with the initial retrieval model and the other tests.

This gain emphasises the quality and quantity of training examples. The error rate obtained in learning phase can also affect the quality of results. When the error rate decreases, the results relevance is improved.

As for Tests 1 and 2, we notice that with a lower number of training data, neural network may not provide W best values, and therefore, the search model may provide null relevance rate. That was the case of queries 3, 4, 5 and 6. Thus, we could conclude that training phase impacted on W values, and in turn on the quality of results provided by our search model.

We can clearly see, with the most queries, that the integration of our new retrieval model provides MAP values more better than the MAP values of our initial retrieval model.

5 CONCLUSIONS

In this paper, we have demonstrated how the neuron network can be used in a vector space model to build a new model of content-based image retrieval. The main idea of this model is to build a connection between the query image and the result score directly via neural network architecture. This model uses neural networks in order to learn an $m \times n$ dimensional matrix W that transforms a m dimensional query vector into a n dimensional score vector whose elements correspond to the similarities of the query vector to the n database vectors.

We have experimentally validated our proposal on 4000 images extracted from caltech 205 collection. The comparative experiments show that our model provides a MAP values better than our initial retrieval model.

The future directions for our work will consist, in first step, in evaluating our new model with a large image collection. In second step, we will use our retrieval model in relevance feedback process.

REFERENCES

Anh, N. D., Bao, P. T., Nam, B. N., and Hoang, N. H. (2010). A new cbir system using sift combined with neural network and graph-based segmentation. In *Proceedings of the Second International Conference on Intelligent Information and Database Systems: Part*

I, ACIIDS'10, pages 294–301, Berlin, Heidelberg. Springer-Verlag.

Arandjelović, R. and Zisserman, A. (2012). Three things everyone should know to improve object retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Berrani, S.-A., Amsaleg, L., and Gros, P. (2002). Recherche par similarité dans les bases de données multidimensionnelles : panorama des techniques d'indexation. *RSTI.*, pages 9–44.

Braquelaire, J. P. and Brun, L. (1997). Comparison and optimization of methods of color image quantization. *IEEE Transactions on Image Processing*, 6(7):1048–1052.

ByoungChul Ko, H. B. (2005). Frip: a region-based image retrieval tool using automatic image segmentation and stepwise boolean and matching. *Trans. Multi.*, pages 105–113.

Caron, Y., Makris, P., and Vincent, N. (2005). Caractérisation d une région d intérêt dans les images. In *Extraction et gestion des connaissances (EGC 2005)*, pages 451–462.

Claveau, V., Tavenard, R., and Amsaleg, L. (2010). Vectorisation des processus d'appariement document-requête. In *CORIA*, pages 313–324. Centre de Publication Universitaire.

Douze, M., Jégou, H., Sandhawalia, H., Amsaleg, L., and Schmid, C. (2009). Evaluation of gist descriptors for web-scale image search. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '09, pages 19:1–19:8, New York, NY, USA. ACM.

Flickner, M. (1997). Query by image and video content: The QBIC system. In *Intelligent Multimedia Information Retrieval*, pages 7–22. American Association for Artificial Intelligence.

Fournier, J., Cord, M., and Philipp-Foliguet, S. (2001). Retin: A content-based image indexing and retrieval system.

Karamti, H. (2013). Vectorisation du modèle d'appariement pour la recherche d'images par le contenu. In *CORIA*, pages 335–340.

Karamti, H., Tmar, M., and BenAmmar, A. (2012). A new relevance feedback approach for multimedia retrieval. In *IKE, July 16-19, Las Vegas Nevada, USA*, page 129.

Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2007). Object retrieval with large vocabularies and fast spatial matching. In *CVPR*.

Philbin, J., Isard, M., Sivic, J., and Zisserman, A. (2010). Descriptor learning for efficient retrieval. In *Proceedings of the 11th European Conference on Computer Vision Conference on Computer Vision: Part III*, ECCV'10, pages 677–691, Berlin, Heidelberg. Springer-Verlag.

Rivero-Moreno, C. and Bres, S. (2003). Les filtres de Hermite et de Gabor donnent-ils des modéles équivalents du systéme visuel humain? In *ORASIS*, pages 423–432.

Schettini, R., Ciocca, G., and Gagliardi, I. (2009). Feature extraction for content-based image retrieval. In *Encyclopedia of Database Systems*, pages 1115–1119.

Srinivasa, K. G., Sridharan, K., Shenoy, P. D., Venugopal, K. R., and Patnaik, L. M. (2006). A neural network based cbir system using sti features and relevance feedback. *Intell. Data Anal.*, 10(2):121–137.

Tollari, S. (2006). *Indexation et recherche d'images par fusion d'informations textuelles et visuelles*. PhD thesis.

Winder, S. A. J., Hua, G., and Brown, M. (2009). Picking the best daisy. In *CVPR*, pages 178–185.

Zhu, Y., Liu, X., and Mio, W. (2007). Content-based image categorization and retrieval using neural networks. In *ICME*, pages 528–531. IEEE.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS