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 information, 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.

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

\[
I_i = \begin{pmatrix} C_{i1} \\ C_{i2} \\ \vdots \\ C_{im} \end{pmatrix} \quad Q_q = \begin{pmatrix} C_{q1} \\ C_{q2} \\ \vdots \\ C_{qm} \end{pmatrix}
\]

The image retrieval process provides a score vector \( S_q = (S_{q1}, S_{q2}, \ldots, S_{qm}) \), 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 \( Q \) associated with a score vector \( S \), we have:

\[
Q \ast W = S
\]

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).

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

\( L_S = \{n_{s1}, n_{s2}, \ldots, n_{sm}\} \) is the output layer, where each neuron \( n_{sj} \) represents a score \( S_{kj} \).

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:

\[
W[i, j] = w_{ij} \quad \forall (i, j) \in \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}
\]

\[
W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mm} \end{pmatrix}
\]

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

For each query \( Q_i = (C_{i1}, C_{i2}, \ldots, C_{im}) \), we propagate \( C_{qi} \) values through the neural network in order to compute \( S_{qi} \) 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,

\( \hat{s}_j \): expected score,

\( \alpha \): learning parameter coefficient,

\( \delta \): error rate.
Algorithm 1: The $w_{ij}$ calculating.

$$\forall (i, j) \in \{1, 2, ... m\} \times \{1, 2, ... n\}$$

$w_{ij} = 1$

for each $Q_j = (c_{q1}, c_{q2}, ..., c_{qm})$

for each $I_i = (c_{i1}, c_{i2}, ..., c_{im})$

$$s_j = \sum_{j=1}^{n} w_{ij} c_{ij}$$

$$\delta_j = s_j - S_j$$

$$w_{ij} \leftarrow w_{ij} + \alpha \delta_j c_{ij}$$

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 \(^3\).

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.
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


