Content-based Image Retrieval System with Relevance Feedback

Hanen Karamti, Mohamed Tmar and Faiez Gargouri
MIRACL, University of Sfax, City Ons, B.P. 242, 3021, Sfax, Tunisia

Keywords: CBIR, Image, Query, Relevance Feedback, Rocchio, Neural Network.

Abstract: In the Content-based image retrieval (CBIR) system, user can express his interest with an image to search images from large database. The retrieval technique uses only the visual contents of images. In recent years with the technological advances, there remain many challenging research problems that continue to attract researchers from multiple disciplines such as the indexing, storing and browsing in the large database. However, traditional methods of image retrieval might not be sufficiently effective when dealing these research problems. Therefore there is a need for an efficient way for facilitate to user to find his need in these large collections of images. Therefore, building a new system to retrieve images using the relevance feedback's technique is necessary in order to deal with such problem of image retrieval.

In this paper, a new CBIR system is proposed to retrieve the similar images by integrating a relevance feedback. This system can be exploited to discover a new proper query representation and to improve the relevance of the retrieved results. The results obtained by our system are illustrated through some experiments on images from the MediaEval2014 collection.

1 INTRODUCTION

The Content-based images retrieval (CBIR), technique to search for images not by keywords but by comparing features of the images themselves, has been the focus of much research and it has gained more attention recently. Consider for instance adding CBIR to different systems of images retrieval such as Google Images, where we would be able to search for images similar to a query image instead of using keywords.

In CBIR, images are indexed by their visual contents such as color, texture, etc. Many research efforts have been made to extract these low level features (Ilbeygi and Shah-Hosseini, 2012), evaluate distance metrics (Tomasev and Mladenic, 2013) and look for efficient searching models (Squire et al., 1999), (Swets and Weng, 1999).

The diversification of search results is increasingly becoming an important topic in the area of images retrieval. Traditional image retrieval systems allows rank the images by their similarity to the query, and relevant images can appear at separate places in the list. Often the user would like to have the relevant images in the first of the list. So, the relevance feedback techniques are appeared as solution to improve the search of such CBIR system.

The concept of relevance feedback (RF) was initially developed, in information retrieval systems (SRI), to improve document retrieval (Salton, 1971). This same technique can be applied to image retrieval. But, Relevance feedback is nonetheless still very well suited for this task as. Most existing RF algorithms are based on techniques for expanding or reformulating query. Automatic query expansion is an effective technique commonly used to add images, from the result, to a query (Rahman et al., 2011). Unfortunately, casual users seldom provide a system with the relevance judgements needed in relevance feedback. In such situation, pseudo-relevance feedback (Wang et al., 2008)(Yan et al., 2003) is commonly used to expand the user-query, where actual input from the user is not required. In this method, a small set of documents are assumed to be relevant without any intervention by the user.

The query expansion method has been actively studied in CBIR systems by using the Standard Rocchio's Formula (Joachims, 1997). Given an image query, the algorithm first retrieves a set of images from the returned result, it combines them with the query to build a new query. While the implementation of the Rocchio formula requires a vector space model to integrate the relevance feedback information. The objective of this paper is to find ways to apply Rocchio formula in CBIR system. For that, we try to design a new vector space model.

Karamti H., Tmar M. and Gargouri F..
Content-based Image Retrieval System with Relevance Feedback.
DOI: 10.5220/0005488502870292
In *Proceedings of the 11th International Conference on Web Information Systems and Technologies* (WEBIST-2015), pages 287-292
ISBN: 978-989-758-106-9

The rest of this paper is organized as follows. Section 2 reviews certain CBIR systems. Section 3 introduces our proposed system. Section 4 reports our experimental results on automatic image search. Section 5 concludes this paper.

2 RELATED WORK

As the network and development of multimedia technologies are becoming more popular, users are not satisfied with the traditional information retrieval techniques. These techniques relies only on low-level features, also known as descriptors. In recent years, a variety of systems have been developed to improve the performance of CBIR, such as the system elaborate by lowe that proposes the scale-invariant feature transform (SIFT) to capture local image information SIFT (Anh,). The SIFT feature detects the salient regions in each image then it describes each local region with a 128 feature vector (Lowe, 2004). Its advantage is that both spatial and appearance information of each local region are recorded in correspondence with the spatial invariance changes in objects. As a result, an image can be viewed as a bag-of-feature-points (BOF) and any object within the image is a subset of the points. With each representation of the (BOF-points), image retrieval is carried out by comparing all feature points in query image to those from all images in the database. Therefore, image retrieval based on BOF-points representation is a solution to the problems with a large database images. Others works are based on the RootSIFT (Arandjelovic, 2012) and the GIST (Douze et al., 2009) that learn the better visual descriptors (than SIFT) (Winder et al., 2009) or better metrics for descriptor comparison and quantization (Daniilidis et al., 2010). 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. Some studies were put forward to change their search spaces, such as color space variation descriptor (pierre 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). 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 introduce a relevance feedback technique (RF) to mend this problem. RF is used in CBIR as is used in text retrieval (Mitra et al., 1998). It change the inital query by a new other query. This new query is build in function with the top of images ranked returned for query retrieval (pseudo-relevance feedback technique). The RF is used to increase the accuracy of image search process. It was first proposed by Rui et. al as an interactive tool in content-based image retrieval (Rui et al., 1998).

In this paper, we propose a new relevance feedback method that integrates 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. This method applies by adapting the standard rocchio formula.

3 PROPOSED SYSTEM

This section describes the architecture of our proposed CBIR system. Given a set of images $(I_1, I_2 ... I_n)$, we aim to build a system that can automatically find images similar to a query image q. Figure 2 shows our system architecture.
The retrieving process is composed of 4 phases:

Indexing ((1), (2)): each image in database is indexed by three descriptors to extract features that describe the image content **(1)**:

- Color layout descriptor (CLD)[1]

[1] CLD is a color descriptor, which 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.

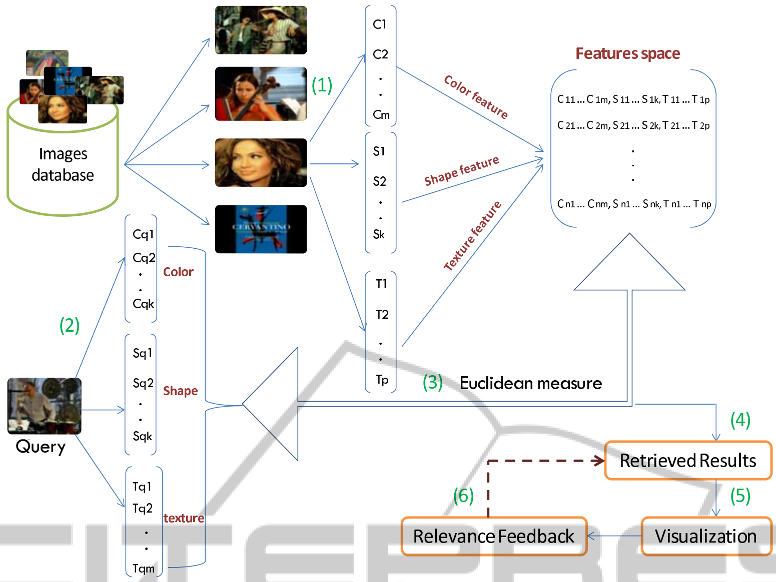

Figure 1: Our CBIR system architecture.

- Gray Level Coocurrence Matrix (GLCM)[2]
- Edge Histogram Descriptor (EHD)[3].

All the visual features are extracted off-line and stored in index which represents the Features space. Each image I is described by a visual descriptor V_i, where V_i is represented by a set of features $(C_{I_1}, C_{I_2} \ldots C_{I_m})$. The same thing for the query **(2)**. The feaures are extracted by the same descriptors to build the visual descriptor $V_q = (C_{q_1}, C_{q_2} \ldots C_{q_m})$.

Retrieving ((3), (4)): To calculate the similarity between representations content of query and images, the system uses the euclidien measure (equation 1) in order to produce for each image a relevance score **(3)**.

Then, results are returned in descending order of relevance **(4)** into a score vector $S_q = (S_{q_1}, S_{q_2} \ldots S_{q_n})$, where n is the number of images in dataset.

$$S_q = dist_{Euclidean}(V_q, V_I) = \sqrt{\sum_{i=1}^{m} (C_{q_i} - C_{I_i})^2} \quad (1)$$

Visualization (5): The purpose of proposed interface's visualization is to visualize the retrieved images for augmenting a user's perception so as communicate him with the interface to improve the search result.

Relevance Feedback (6): For the relevance feedback phase, system recovers the more relevant images and it rebuilds a new query, in taking into account the selected images. This is done by adapting the Standard Rocchio's Formula (Joachims, 1997).

The objective of this paper is to find ways to apply Rocchio in a CBIR system. First, we try to design a vector space model: this transformation is called vectorization. There is a model of vectorization in information retrieval, our contribution is to develop a model of vectorization in CBIR.

3.1 Vectorization Method

Conversion of the feature vector of an image to a vector score form is performed with the neural network. The network is constructed by two layers:

- input layer: corresponds to the features values of query (V_q).
- output layer: corresponds to the score values obtained by initial search of query (S_q).

Given a set of queries $(q_1, q_2, \ldots, q_n)$, where each query is represented by a features vector and a scores vector, we put them into our neuron network by propagating feature values from a set of score values.

For each query q_i represented by V_i, we propagate C_{q_i} values through the neural network in order to compute S_{q_j} scores 2.

$$V_q * W = S_q \quad (2)$$

[2]GLCM: method is a way of extracting second order statistical texture features. The approach has been used in a number of applications. A GLCM is a matrix where the number of rows and colums is equal to the number of gray levels in the image.

[3]EHD is a shape descriptor, which is proposed for MPEG-7 expresses only the local edge distribution in the image.

Each connecting line, between input layer and output layer, has an associated weight w_{ij}. Our neural network is trained by adjusting these input weights (connection weights), so that the calculated outputs may be approximated by the desired values. For propagation between features vector and scores vector, we adapt the calculating weight algorithm (Karamti et al., 2014).

$$W[i,j] = w_{ij} \qquad \forall (i,j) \in \{1,2...m\} \times \{1,2...n\} \quad (3)$$

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

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. Where:

Algorithm 1: Propagation algorithm.

$\forall (i,j) \in \{1,2...n\}\{1,2...m\}$
$w_{ij} = 1$
for *each* query q_i **do**
  for $eachF_{qi} = (F_{qi1}, F_{qi2}...F_{qim})$ **do**
    $S_{qi} = (S_{qi1}, S_{qi2}...S_{qim})$
    $s_j = \sum_{j=1}^{m} w_{ij} F_{ij}$
    $w_{ij} = w_{ij} + \alpha F_{ij}$
  end for
end for

s_j is the actual score, and α is the learning parameter coefficient ($0 \leq alpha < 1$).

3.2 Integration of Standard Rocchio

To improve the results found by vectorization technique, we integrate the Standard Rocchio formula 4. The Standard Rocchio's Formula is coming from the documentary information retrieval. This method requires that the user selects from the displayed images some relevant images and some non-relevant ones. In this paper, the relevant images are automatically selected following the pseudo-relevant feedback technique. This method assume that the first retrieved images k are relevant and uses these images to build the new query.

Rocchio is given by the following formula:

$$q_{new} = q_{old} + \frac{1}{|R|} \sum_{i \in R} i - \frac{1}{|S|} \sum_{i \in S} i \qquad (4)$$

where:
- q_{old}: the query issued by the user.
- q_{new}: the new query.
- R: relevant images.
- S: irrelevant images.
- i: an image of dataset.

4 EXPERIMENTAL SETUP

We conduct experiments to evaluate the impact of vectorization and rocchio's adaptation on predicting the best relevance feedback model associated with a query. We compare the performance of our best retrieval system with other CBIR systems.

4.1 Data Set

For the retrieval experiments we rely on the images corpus from MediaEal 2014 (Ionescu et al., 2014). The mediaEval2014 data set consists of 300 locations (e.g., monuments, cathedrals, bridges, sites, etc) spread over 35 countries around the world. Data is divided into a development set called devset, containing 30 locations, intended for designing the approaches. A test set called testset, containing 123 locations, to be used for the official evaluation. All the data was retrieved from Flickr (devset fournit 8923 images et TestSet 36452).

4.2 Evaluation

Performance is assessed for both diversity and relevance. The following metrics are computed:

- Cluster Recall at X ($CR@X$): a measure that assesses how many different clusters from the ground truth are represented among the top X results (only relevant images are considered).
- Precision at X ($P@X$): measures the number of relevant photos among the top X results and
- F1-measure at X ($F1@X$): is the harmonic mean of the previous two. Various cut off points are to be considered, i.e., $X = 5, 10, 20, 30, 40, 50$.

Official ranking metric is the $F1@20$ which gives equal importance to diversity (via $CR@20$) and relevance (via $P@20$). This metric simulates the content of a CBIR system and reflects his behavior relative to other systems.

4.3 Results

Figure 2 shows the values of average $P@20$, $CR@20$ and $F1@20$ results for the Run1. It's a simple example to show that the accuracy variation in function of Cluster Recall and F1-measure. What is in-

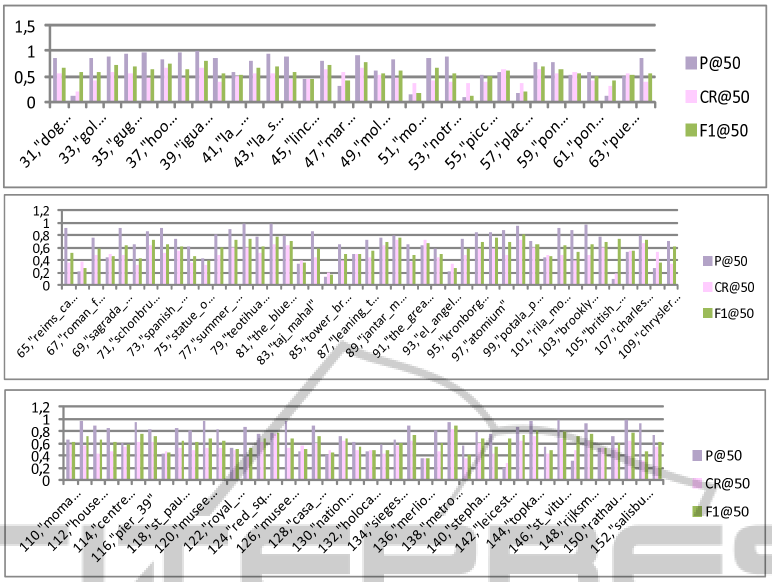

Figure 2: Evaluation of the 153 queries with $P@50$, $CR@50$ and $F1@50$.

teresting to observe is the fact that the highest precision is achieved with the majority of queries. While the results for the CBIR comparison methods on MediaEva2014 collections are presented in the following table:

Table 1: Retrieval Performances between initial search (Run 1), vectorization method (Run 2) and Rocchio method (Run 3).

Run name	$P@20$	$CR@20$	$F1@20$
Run 1	0.7772	0.3265	0.4501
Run 2	0.7516	0.314	0.4329
Run 3	**0.7879**	**0.4426**	**0.5552**

- Run 1: is our initial retrieving system.
- Run 2: is our vectorization method.
- Run 3: is our CBIR system by applying our vectorization method and the standard Rocchio formula.

In fact, Table 1 summarises the results when returning the top 20 images per location. We notice that the found results by using the vectorization process are very close to the initial system results.

Our goal by the application of vectorization method is to find results that are very close to the results by initial retrieval system (Run 1). Indeed, the results found by the vectorization method (Run 2) show that the transformation of the matching process to a vector space model does not cause of informations loss.

Once the research model is vectorized, we can added the standard formula Rocchio. For this adaptation, we should choose a k number of the top retrieved images (k is fixed to 50). Then, we can build a new query in function of k and the old query used in initial search. The proved results by adapting the Standard Rocchio Formula (Run 3). With table 1, we can notice that our RF method, based on Rocchio Formula (Run 3), gives better results compared to the two others runs.

5 CONCLUSIONS

In this paper, we presented a new system of CBIR that contains a new technique of relevance feedback based on the Sandard Rocchio Formula. This technique needs a transformation from a matching model based on vectors of features to a matching model based on vectors of scores. In addition, we have shown how a vectorization model can be used to enhance the retrieval accuracy. Finally, we compare the performance of the proposed system with three performance measures.

Experimental results show that it is more effective and efficient to retrieve visually similar images with a relevance feedback technique based on scores's vector representation of images.

REFERENCES

Arandjelovic, R. (2012). Three things everyone should know to improve object retrieval. In *Proceedings of* *the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, pages 2911–2918, Washington, DC, USA. IEEE Computer Society.

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.

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

Daniilidis, K., Maragos, P., and Paragios, N., editors (2010). *ECCV'10: Proceedings of the 11th European Conference on Computer Vision Conference on Computer Vision: Part III*, Berlin, Heidelberg. Springer-Verlag.

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.

Ilbeygi, M. and Shah-Hosseini, H. (2012). A novel fuzzy facial expression recognition system based on facial feature extraction from color face images. *Eng. Appl. Artif. Intell.*, pages 130–146.

Ionescu, B., Popescu, A., Lupu, M., Gînsca, A., and Müller, H. (2014). Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. In *Working Notes Proceedings of the MediaEval 2014 Workshop, Barcelona, Catalunya, Spain, October 16-17, 2014.*

Joachims, T. (1997). A probabilistic analysis of the rocchio algorithm with tfidf for text categorization. In *Proceedings of the Fourteenth International Conference on Machine Learning*, ICML '97, pages 143–151, San Francisco, CA, USA.

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.

Karamti, H., Tmar, M., and Gargouri, F. (2014). Vectorization of content-based image retrieval process using neural network. In *ICEIS 2014 - Proceedings of the 16th International Conference on Enterprise Information Systems, Volume 2, Lisbon, Portugal, 27-30 April, 2014*, pages 435–439.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110.

Mitra, M., Singhal, A., and Buckley, C. (1998). Improving automatic query expansion. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 206–214, New York, NY, USA. ACM.

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

pierre Braquelaire, J. and Brun, L. (1997). Comparison and optimization of methods of color image quantization. *IEEE TRANS. ON IMAGE PROCESSING*, 6:1048–1051.

Rahman, M. M., Antani, S. K., and Thoma, G. R. (2011). A query expansion framework in image retrieval domain based on local and global analysis. *Inf. Process. Manage.*, 47(5):676–691.

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.

Rui, Y., Huang, T. S., Ortega, M., and Mehrotra, S. (1998). Relevance feedback: A power tool for interactive content-based image retrieval.

Salton, G. (1971). *The SMART Retrieval System—Experiments in Automatic Document Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Squire, D. M., Müller, H., and Müller, W. (1999). Improving response time by search pruning in a content-based image retrieval system, using inverted file techniques. In *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, CBAIVL '99, pages 45–, Washington, DC, USA. IEEE Computer Society.

Swets, D. L. and Weng, J. (1999). Hierarchical discriminant analysis for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21:386–401.

Tomasev, N. and Mladenic, D. (2013). Image hub explorer: Evaluating representations and metrics for content-based image retrieval and object recognition. In Blockeel, H., Kersting, K., Nijssen, S., and Zelezn, F., editors, *ECML/PKDD (3)*, volume 8190, pages 637–640. Springer.

Wang, X., Fang, H., and Zhai, C. (2008). A study of methods for negative relevance feedback. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 219–226, New York, NY, USA. ACM.

Winder, S. A. J., Hua, G., and Brown, M. (2009). Picking the best DAISY. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*, pages 178–185.

Yan, R., Hauptmann, A. G., and Jin, R. (2003). Negative pseudo-relevance feedback in content-based video retrieval. In *In Proceedings of ACM Multimedia (MM2003*, pages 343–346.