

# MIRACL's participation at MediaEval 2014 Retrieving Diverse Social Images Task

Hanen Karamti  
MIRACL Laboratory  
City ons, University of Sfax  
B.P.3021 Sfax TUNISIA  
karamti.hanen@gmail.com

Mohamed Tmar  
MIRACL Laboratory  
City ons, University of Sfax  
B.P.3021 Sfax TUNISIA  
mohamedtmar@yahoo.fr

Faiez Gargouri  
MIRACL Laboratory  
City ons, University of Sfax  
B.P.3021 Sfax TUNISIA  
faiez.gargouri@fsegs.rnu.tn

## ABSTRACT

The 2014 MediaEval Retrieving Diverse Social Images Task tackles the problem of search result diversification of Flickr results sets formed from queries about geographic places and landmarks. In this paper we describe our approach using a neuron network. This approach uses only the visual information from image. The goal of this method is to 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.

We submitted two runs. The first run describes a sample method of content-based image retrieval. The second run describes our approach of query expansion using our new retrieval model. Our two runs produced a high precision of the results.

## 1. INTRODUCTION

The diversification of search results is increasingly becoming an important topic in the area of information retrieval. The MediaEval 2014 Retrieving Diverse Social Images Task addresses the problem of result diversification in the context of social photo retrieval [1]. The data consist of a development set containing 30 locations (devset), a user annotation credibility set containing information for ca. 300 locations and 685 users (credibilityset) and a test set containing 123 locations (testset). Data was retrieved from Flickr using the name of the location as query. Participants will receive for each location a list of 300 photos retrieved from and ranked with Flickr's default relevance algorithm.

For each query, our strategy to induce diversity while keeping the relevance is based on four steps:

- Step 1: Recalculate the scores of the provided ranked list using feature extraction;
- Step 2: Rerank the result to improve relevance;
- Step 3: Use the results (the score vector and the feature vector for each image) to construct a new retrieval model using the neuron network;
- Step 4: Calculate the new scores of images using our new model of search;
- Step 5: Finally rerank the results in descending order of their scores.

The goal here is to produce a ranked list of images that are both relevant and diverse in response to a location-based query. This list ranks up to 50 photos.

Throughout our work, we aimed to maximise the  $CR@20$  score. Below, we describe our methods for the two runs and discussing the results.

## 2. APPROACHES

To obtain representative and diverse photos in the upper rank, we use a method based on the visual information extracted from images. Run 1 uses a sample method of feature extraction [2] and Run 5 uses a new vectorization method using a neuron network [3]. In this section, we explain both methods and features briefly.

### 2.1 Run 1: Extraction of Visual features

Data extraction processing and query processing are two main functionalities supported of feature extraction [3]:

The data extraction process is responsible for extracting appropriate features from images and storing them into their feature vectors. This process is usually performed offline. The architecture of this phase is described in Figure 1. A feature vector  $V_I$  of an image  $I$  can be thought of as a list of low-levels features  $(C_1, C_2, \dots, C_m)$ , where  $m$  is the number of features. We have used three descriptors: A color layout descriptor (CLD)<sup>1</sup>, an Edge Histogram Descriptor (EHD)<sup>2</sup>, and a Scalable Color Descriptor (SCD)<sup>3</sup>. The  $C_i$  represents the combination of  $C_{CLD}$ ,  $C_{SCD}$  and  $C_{EHD}$  of feature  $i$ .

The query processing, in turn, extracts a feature vector from a query and applies a metric (Euclidean distance, see equation 1) to evaluate the similarity between the query image and the database images.

$$dist_{Euclidean}(V_I, V_{I_i}) = \sqrt{\sum_{i=1}^m (C_{I_j} - C_{I_{ij}})^2} \quad (1)$$

where  $V_I$  is the feature vector of an image  $I$ ,  $V_{I_i}$  is the feature vector of an image  $I_i$ ,  $C_{I_j}$  is the low-level feature

<sup>1</sup>is designed to capture the spatial distribution of color in an image.

<sup>2</sup>is a texture descriptor proposed for MPEG-7 and expresses only the local edge distribution in the image.

<sup>3</sup>the histogram is generated by color quantizing the image into 256 bins in the HSV color space, with 16 bins for hue, and 4 bins each for saturation and value.

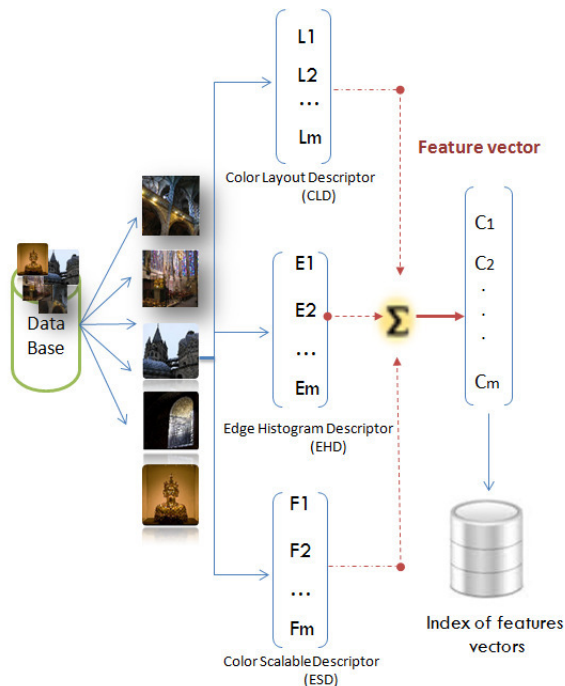


Figure 1: Architecture of data extraction processing

$C_{I_j}$  corresponding to  $V_I$  and  $C_{I_{ij}}$  is the low-level feature  $C_{I_{ij}}$  corresponding to  $V_{ij}$ .

The similarities scores of the queries results builds a score vector. A score vector  $S_I$  of an image  $I$  can be thought of as a set of scores  $(S_1, S_2, \dots, S_n)$ , where  $n$  is the dimension of database images.

## 2.2 Run 5: vectorization process with the visual features

Each query location  $q_i$  is expressed on the feature vector by  $(F_{q_{i1}}, F_{q_{i2}}, \dots, F_{q_{im}})$ , where  $F_{q_{ij}}$  is the value of feature in the query  $q_i$ . The feature vector  $F_{q_i}$  corresponding to  $q_i$  is computed with the same method of a score vector  $S_I$ .

The image retrieval process provides a score vector  $(S_{q_{i1}}, S_{q_{i2}}, \dots, S_{q_{in}})$ , where  $S_{q_{ij}}$  is the value of similarity score between the query  $q_i$  and the image  $j$  where  $j \in \{1, \dots, n\}$ .  $S_{q_i}$  represents the list of 300 photos retrieved from Flickr and ranked with query processing process described in previous section.

This vector space model is characterized by a  $(m \times n)$  matrix  $W$  where for each query  $q_i$ , described by the  $F_{q_i}$  feature vector, associated with a score vector  $S_{q_i}$ , we have:

$$F_{q_i} \times W = S_{q_i} \quad (2)$$

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

$$\begin{pmatrix} F_{q_{i1}} \\ F_{q_{i2}} \\ \vdots \\ F_{q_{im}} \end{pmatrix} \times \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nm} \end{pmatrix} \cong \begin{pmatrix} S_{q_{i1}} \\ S_{q_{i2}} \\ \vdots \\ S_{q_{in}} \end{pmatrix} \quad (4)$$

The  $w_{ij}$  values are calculated with a propagate algorithm using a neural network [3].

Table 1: Results for all runs in MediaEval 2014

Run name	$P@20$	$CR@20$	$F1@20$
Run 1	0.7772	0.3265	0.4501
Run 5	0.7516	0.314	0.4329

## 3. RESULTS

Performance is going to be assessed for both diversity and relevance. The performance of our approach is computed by:

- Average  $P@20$ : Precision, a measure that assesses the number of relevant photos among the top 20 results;
- Average  $CR@20$ : Cluster Recall, a measure that assesses how many different clusters from the ground truth are represented among the top 20 results (only relevant images are considered);
- Average  $F1@20$ :  $F1$ -measure at 20 is the harmonic mean of the previous two.

Table 1 shows the evaluated results of our two submission runs. We have submitted for evaluation two runs, employing only the visual descriptors mentioned in previous section: *Run 1* is the basic method of feature extraction and *Run 5* corresponds to the vectorization process.

In fact, Table 1 summarises the results when returning the top 20 images per location. We notice that the visual features provides a high precision of the results. A slight improvement in precision came in use of neuron network to produce a new image retrieval model. Our results are close for the two runs, despite the variety of algorithms used.

## 4. CONCLUSIONS

In this paper we present our first participation in MediaEval 2014. Our results show that a purely visual approach can lead to efficient results. However, improved results could be obtained by leveraging on other types of information (textual description), if available, for further refining of the results. In the future, we plan to explore the integration of social and visual cues in order to obtain a more efficient diversification.

## 5. REFERENCES

- [1] B. Ionescu, A. Popescu, M. Lupu, A. Gînscă, and H. Müller. Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. In *Proceedings of the MediaEval 2014 Multimedia Benchmark Workshop, Barcelona, Spain, October 16-17, 2014*.
- [2] H. Karamti, S. Fakhfakh, M. Tmar, and F. Gargouri. MIRACL at imageclef 2014: Plant identification task. In *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18*, pages 747–755, 2014.
- [3] H. Karamti, M. Tmar, and F. Gargouri. 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*, pages 435–439, 2014.