Combining Features by Query-time Weights Determination for Image Retrieval

Nikolay Neshov[1]

Abstract – **In this paper a Content Based Image Retrieval approach is presented that combines weighted features. The determination of feature's weight involves examination by *K* additional queries formed by the first *K* retrieved images. For each additional query, the user query is added to the database and the standard deviation of its rank is utilized to calculate the feature's weight. Finally it is shown the performance of the system.**

Keywords – **Content Based Image Retrieval (CBIR), Late Fusion, Weights Determination, Lucene Image Retrieval (LIRe)**

I. INTRODUCTION

The necessity of CBIR system with better performance is growing due to the increased number of digital images in the last years. Although there are a lot of implementations of CBIR systems [1], they still have a limited accuracy. This is mainly because of the fact that the semantic content of images is not described appropriately by using only their low-level features. The database images that depict natural photos can be described more accurate by utilizing different type of features, such as color, texture and shapes. In order to increase the performance of the system, the challenge is to find out a way to fuse the information from all available visual features when measure the similarity between each image from the database and user query. There are two main approaches for fusion: early fusion, where multiple image feature vectors are combined to form a new feature vector, and late fusion, where the result lists produced by the distinct searches (using each available feature) are combined in a new ranked list. There has been much work done in the area of combination of features for CBIR. Generally speaking, the late fusion methods are most widely used and developed [2]. For comparison purposes, in this paper we have implemented the following five fusion methods: CombSUM with Min-Max normalization (CombSUM Min-Max), CombSUM with Z-Score normalization (CombSUM Z-Score), Borda Count and Inverse Ranking Position [4]. All of these methods use equal weights of the scores (or ranks) to produce the final score (or rank) for each image from the database. In contrast, the method suggested in the paper is based on associating weights to the features with respect to the content of the query. The calculation of feature's weight is done by testing the system by *K* additional queries that are formed by the first *K* retrieved images. For each additional query, the initial query is added to the database and the standard deviation of its rank is used to calculate the feature's weight. Finally a weighted sum of the scores produced by each feature is utilized to compute the similarity score between initial query and each image in database. The algorithm suggested in this work is implemented with the help of Lucene Image Retrieval (LIRe) [5], [6] - an open source Java library. The comparison test to other approaches such as CombSUM Min-Max, CombSUM Z-Score, Borda Count and Inverse Ranking Position show that system based on the proposed method have a higher accuracy.

[1]Nikolay Neshov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: nneshov@tu-sofia.bg.

II. DATABASE AND PERFORMANCE MEASURES

For experiments of the proposed algorithm, we used WANG Database [7]. It contains a total of 1000 images manually divided in 10 classes of 100 images each. To estimate the retrieval accuracy of the system each image from the database is used as query. Then by scanning the list of results if the current image belongs to the query's category it is considered relevant. The performance measure utilized is Mean Average Precision (MAP) [8].

III. FEATURES CONSIDERED FOR COMBINATION

Based on the results provided by M. Lux [6] for the WANG database, where the retrieval performance of 11 descriptors has been investigated, we selected the top four of them (those that have the highest accuracy in terms of MAP). The features extracted by these descriptors are considered for combination. Table I. summarizes the details for each of descriptor (MAP and dimension of the respective feature vector).

TABLE I
MAP FOR ALL FOUR DESCRIPTORS CONSIDERED FOR COMBINATION AND DIMENSION OF RESPECTIVE FEATURE VECTORS

Descriptor	MAP, %	Feature vector dimension
Joint Composite Descriptor (JCD)	50,95	168
Color Histogram (CH)	48,44	512
Correlogram Autocorrelogram (CA)	47,51	1024
DCT Histogram (DCTH)	44,55	192

A brief explanation for all descriptors used as a background in this paper can be found in [9].

IV. ALGORITHM DESCRIPTION

In order to explain the idea of proposed algorithm we consider the following: Suppose we have a database (DB) that

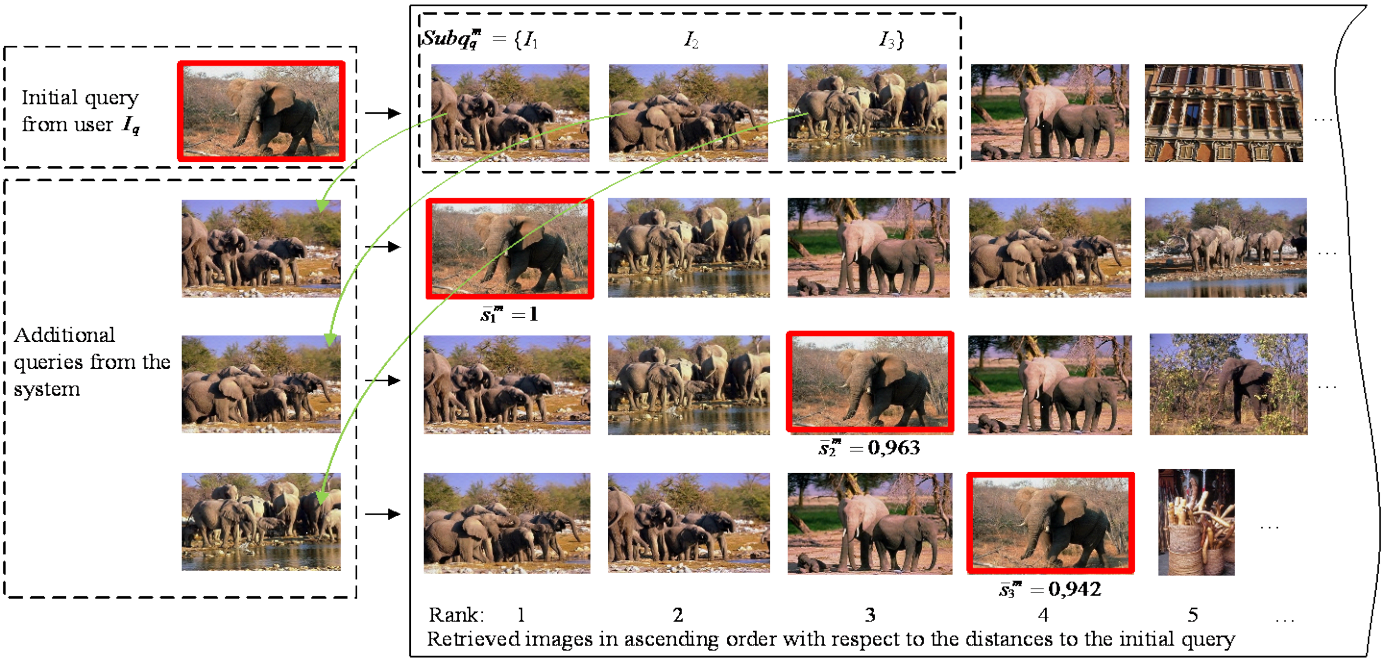

Fig. 1. Example describing additional requests formulation for feature's weight determination

contains N images and CBIR system in which M descriptors are implemented to extract their visual features. When the user is searching for a query I_q, the system extracts m-th feature vector (m=[1;M]) and calculates the distances to all feature vectors of the images in database. An initial retrieval list L_q^m is then generated that contains R images from the database in ascending order of distances to the query. That is $L_q^m = \{I_1, ..., I_r, ..., I_R\}$, where I_r is the r-th retrieved image. Then the system (using descriptor m) is examined by K additional queries, in which the query I_k (k=[1;K]) is the k-th retrieved image from the beginning of L_q^m. Let's denote the subset of additional queries for initial query q as $Subq_q^m = \{I_1, ..., I_k, ..., I_K\}$, $Subq_q^m \subset L_q^m$, and $k \equiv r$, because $K < R$. Furthermore each k-th request is carry out by adding I_q to DB. This kind of query formulation is based on an assumption that the first K retrieved images are more likely to be relevant to the initial query. Hence I_q should be found in the top ranks of resulted list, when the system is requested by I_k. A visual example of requests, structured in this way (in case of K = 3) can be seen in Fig. 1. Given image I_q to the input of the CBIR system, it retrieves R images as result. Next, three additional requests are performed using the first 3 retrieved images $\{I_1, I_2, I_3\}$. It can be seen (in red rectangles) that the initial query I_q is retrieved in the top ranking positions of all three lists produced as result of the additional queries. Moreover the rank of I_q increases as k increases. The observed effect is expected because the more dissimilar image (with respect to I_q) is given to the system's input, the more distant position for the initial query should be judged. The method developed in this work is based on a hypothesis that accuracy of the system (hence the performance of the descriptor m) becomes higher if the effect described above is more pronounced.

In the classical approach of CBIR, the results are only based on the similarity between I_q and images in DB. The suggested approach for additional query formulation takes into account also the similarity between each k-th image and images in the database. For the example in Fig. 1 it can be seen that when the query is I_q, there is only one image (I_1) which is more similar than I_2. On the other hand if the query is I_2, there are two images from the database - more similar than I_q.

Given initial query I_q, for each descriptor m, (m = [1;M]), M dependencies of the query rank as a function of k (i.e. $r_k^m = f(k)$) can be obtained. The analysis of deviation in the ranks (as k increases) makes it possible to measure the presence of the observed effect in the hypothesis described above. Fig. 2 depicts the functions $f(k)$ for one image from the WANG database (K = 10 and M = 5).

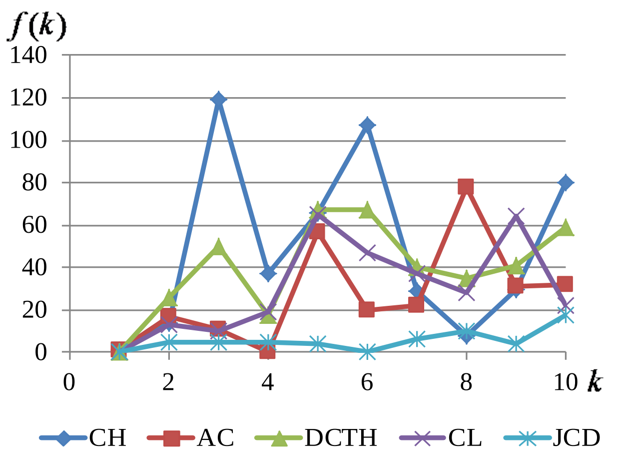

Fig. 2. Dependency of initial query rang from the additional queries for each descriptor and K = 10

It is shown that the deviation of k can be positive as well as negative. Fig. 3 depicts the functions $f(k)$ for the same image when $K = 500$ and step of incensement of k is 10.

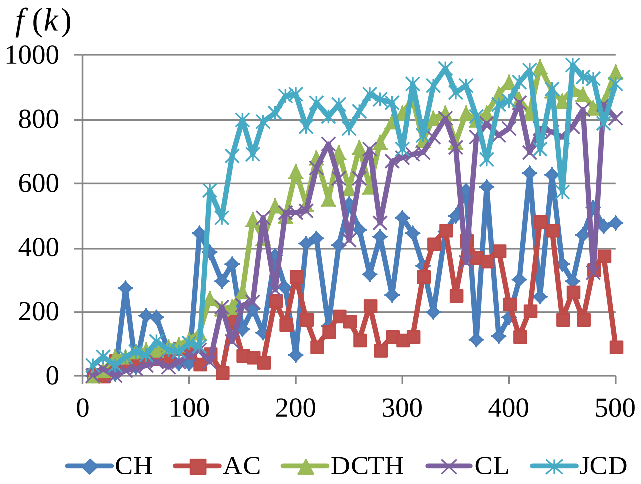

Fig. 3. Dependency of initial query rang from the additional queries for each descriptor and $K = 500$

Despite the fluctuations in $f(k)$ it can be observed that for all descriptors, the functions tend to increase at sufficiently large values of K. According to the foregoing hypothesis, those descriptors for which smaller values of the standard deviation σ_q^m of $f(k)$ are produced should be given greater weights.

The value of σ_q^m is given by:

$$\sigma_q^m = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(r_k^m - \mu_q^m)^2}\,, \qquad (1)$$

where

$$\mu_q^m = \frac{1}{K}\sum_{k=1}^{K} r_k^m\,, \qquad (2)$$

The final similarity score $sim'(I_q, I_n)$ between I_q and image from tha database I_n , n=[1;N] is given as wighted average score:

$$sim'(I_q, I_n) = \frac{1}{M}\sum_{m=1}^{M} p^m.\bar{s}_n^m\,, \qquad (3)$$

where $\bar{s}_n^m$ is normalized similarity score between I_q and I_n which is calculated by using MIN-MAX normalization [2]:

$$\bar{s}_n^m = \frac{s_n^m - s_n^{\min}}{s_n^{\max} - s_n^{\min}}\,, \qquad (4)$$

where $s_n^{\max}$ and $s_n^{\min}$ denote respectively the maximum and the minimum similarity score with respect to the query image and s_n^m is the similarity score, determined from the distance between feature's vectors of I_q and I_n:

$$s_n^m = d_{\max}^m - d_n^m\,. \qquad (5)$$

The feature's weight p^m in Eq. 3 is given by:

$$p^m = \frac{\frac{1}{\sigma_q^m}}{\sum_{m=1}^{M}\frac{1}{\sigma_q^m}}\,, \qquad (6)$$

in which $\sum_{m=1}^{M} p^m = 1$.

For the investigation of the proposed method, the value of $K = 5$. The choice of this value is based on experimental tests which show that if K is greater than 5, the accuracy performance of the system is not increasing any more. Furthermore, any additional query decreases the retrieval performance which is not desirable especially in case of large databases.

V. Experimental Results

Table II shows an experimental comparison of the performance of the proposed combination approach in this article and several late fusion methods available in the literature for different sets of descriptors. It can be seen that proposed algorithm reaches the highest value of MAP among all others examined techniques. The increase in performance in comparison to all other methods is higher for each features' sets and is the highest when all four features are combined.

Table II
Performance comparison in terms of MAP [%] of the proposed method and four other late fusion methods

Descriptors' sets / Fusion method	JCD CH CA	JCD CH DCTH	JCD CA DCTH	CH CA DCTH	**JCD CH CA DCTH**
Proposed Method	**56,9**	**58,13**	**61,5**	**60,66**	**61,58**
CombSUM Min-Max	56,44	57,94	61,18	60,19	60,71
CombSUM Z-Score	56,86	58,04	61,43	60,59	61,09
Borda Count	55,95	56	59,31	58,76	59,71
Inverse Ranking Position	54,15	54,12	55,78	54,97	56,26

Fig. 4 allows for a visual interpretation of the performance values from Table I. It can be noted that the fusion methods based on rank (Borda Count and Inverse Ranking Position) reach lower values of MAP compared to the other methods based on scores.

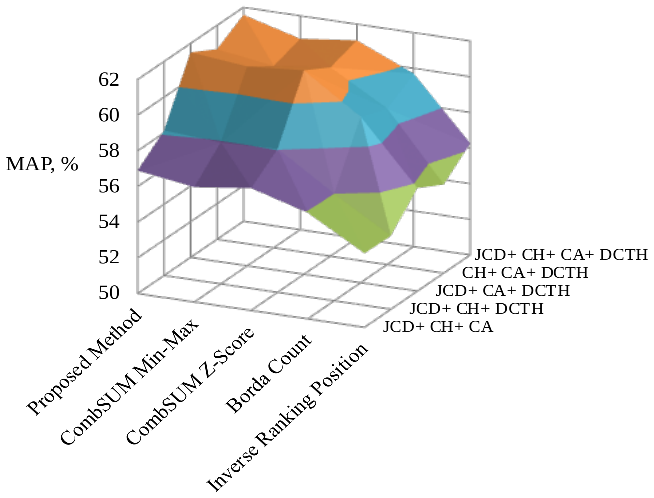

Fig. 4. Graphical comparison of Mean Average Precision, of the proposed method and four other late fusion methods

VI. Conclusions

The contribution of this work has both scientific and practical aspects.

From the scientific point of view this paper suggests an approach for CBIR, which is based on combination of features by query-time weights determination. The method is compared to four well known late fusion methods and the investigations show increasing of retrieval performance in terms of MAP (Table II). Compared to the most accurate baseline descriptor (JCD) the improvement of MAP is from 50,95 % (Table I) to 61,58 % (more than 20 %).

From the practical point of view a system for Content Based Image Retrieval is built that represents an extension of LIRe.

References

[1] M. Danish, P. Rawat, S. Ratika, "Content Based Image Retrieval Based On Color, Texture, Shape & Neuro Fuzzy", Survey, Int. Journal Of Engineering Research And Applications, Vol. 3, No. 5., pp. 839-844, 2013.

[2] A. Depeursinge, H. Muller, *Fusion techniques for combining textual and visual information retrieval*, In: ImageCLEF: Experimental Evaluation in Visual Information Retrieval. Springer, Heidelberg, 2010.

[3] S.A. Chatzichristofis, K. Zagoris, Y. Boutalis, A. Arampatzis, "A Fuzzy Rank-Based Late Fusion Method for Image Retrieval", In: K. Schoeffmann, B. Merialdo, A.G. Hauptmann, C.W. Ngo, Y. Andreopoulos, C. Breiteneder (eds.) MMM 2012. LNCS, vol. 7131, pp. 463–472. Springer, Heidelberg, 2012.

[4] N. Neshov, "Comparison on Late Fusion Methods of Low Level Features for Content Based Image Retrieval", (LNCS) 23rd International Conference on Artificial Neural Networks (ICANN) Sofia, Bulgaria, September 10-13, Vol. 813, No 1, pp. 619-627, Springer-Verlag Berlin Heidelberg, 2013.

[5] M. Lux, S. Chatzichristofis, "LIRe: Lucene Image Retrieval – An Extensible Java CBIR Library" In: Proceedings of the 16th ACM International Conference on Multimedia, Vancouver, Canada, pp. 1085–1088, 2008.

[6] Lux, M.: "Content Based Image Retrieval with LIRE" In: Proceedings of the 19th ACM, International Conference on Multimedia, Scottsdale, Arizona, USA, pp. 735–738, 2011.

[7] J. Z. Wang, J. Li, G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIbraries", In: the IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 9, pp. 947–963, 2001.

[8] A.A. Popova, Neshov, N.N., "Combining Features Evaluation Approach in Content-Based Image Search for Medical Applications" In: Kountchev, R., Iantovics, B. (eds.): Adv. in Intell. Anal. of Med. Data & Decis., vol. 473, no. 1, pp. 113–123, Springer-Verlag Berlin Heidelberg, 2013.

[9] A. A. Popova, N. Neshov, "Image Similarity Search Approach Based On The Best Features Ranking", Egyptian Computer Science Journal, Vol. 37, No. 1, pp. 51–65, 2013.