Web Visual Search Using Re-ranking Method and Neural Network Machine Learning

Milena Vesić, Goran Zajić, Nenad Kojić and Ana Gavrovska, *Member, IEEE*

***Abstract*—There is a continuing growth of online visual data. The latest development of web technologies enabled the efficient access to image and video databases using online services. The visual search is demanding from the efficiency standpoint. In this paper we proposed re-ranking for content based information retrieval improvement. The machine learning based system with relevance feedback uses Euclidean distance based initial search with re-ranking in order to provide a user with relevant images. The relevant images are selected according to feature standard deviation, feature participation importance and feature correlation. Re-ranked images enable more efficient search in the initial phase. User conducts the search using the relevance feedback in the desired direction.**

***Index Terms*—Visual search, web, features, reranking, distance, machine learning.**

I. Introduction

With accelerated web technology development, there is an efficient image and video content exchange over online services. Huge amounts of images are located in distributed image databases around the world, and they are available to users for different purposes (big data). An effective search for large collections of images represents a challenge for some time.

One of the important drivers when developing these technologies is the commercial use of the content based search, in order to develop marketing or purchasing. Direct implementation of these technologies can be found within large web browsers (Google, Bing, Pinterest), and within online store systems (eBay).

Web search engines for images have been very popular. Content based web search for images [1, 2] are based on query image selection which can be recorded by a user (upload) or selected from the available image databases. Low level features are extracted (color, texture, shape), and further used for comparison of image features from the database. A distance metric can be applied as an objective measure of similarity (e.g. Euclidean distance). Search engines can be oriented towards particular features, or towards a feature vector found by their concatenation [3]. It is also possible to perform an efficient search with a part of the feature vector [3].

There is a difference between an objective similarity measure and an intuitive user assessment, which is the result of individual perception. Thus, a relevance feedback as user intervention is applied, as well as intelligent logic for input query correction [2, 3]. The result of initial search, as the result of an objective similarity measure, represents a set of a certain number of images which are presented to the user. The query is corrected according to relevant and not-relevant images selected by the user from the presented set. Relevance feedback and intelligent logic can be applied iteratively.

The success of the relevance feedback depends on the number of relevant images that are presented to the user after the initial search. In order to increase the number of relevant images, a reranking method can be applied. Images similar to the query image, obtained using objective similarity measure, are re-ranked. The re-ranking is performed by a new similarity criterion.

There are different approaches available in the realization of the re-ranking [4]. One of the interesting ways to perform the re-ranking is based on counting the images which appear in clusters (Click-based) [5]. Another approach is to rank images by their appearance in the search results according to different features (automatic ranking) [6].

In this paper, a re-ranking method is proposed. After the initial image search, proposed re-ranking is applied according to three criteria. The first criterion is based on the standard deviation of query image features. The second criterion is the feature participation importance. A percentage of squared distance between particular feature values in total distance between feature vectors of query and analyzed images is calculated. The third criterion for ranking is the correlation of corresponding features of query and analyzed images. The total new image ranking is calculated by the product of three parameters based on the three criteria. The proposed re-ranking is tested using the image database consisted of 1000 images. Re-ranking may affect efficiency in the initial search, globally for the entire image database, and locally for the specific classes. After the initial search, user continues to search in the desired direction.

The paper is organized as follows. After the introduction, a description of the web visual search system is given in Section II. Used image features for image characterization are described in Section III. In addition to the features, this

Milena Vesić is with the ICT College of Vocational Studies in Belgrade, Zdravka Celara 16, 11020 Belgrade, Serbia (e-mail: milena.vesic@ ict.edu.rs).

Goran Zajić is with the ICT College of Vocational Studies in Belgrade, Zdravka Celara 16, 11020 Belgrade, Serbia (e-mail: gzajic@gmail.com, goran.zajic@ict.edu.rs).

Nenad Kojić is with the ICT College of Vocational Studies in Belgrade, Zdravka Celara 16, 11020 Belgrade, Serbia (e-mail: nenad.kojic@ict.edu.rs).

Ana Gavrovska is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia (e-mail: anaga777@etf.bg.ac.rs, anaga777@gmail.com)..

Section also describes the proposed re-ranking method. The obtained test results using the proposed re-ranking are presented in Section IV. This is followed by appropriate discussion. Section V is dedicated to the conclusions.

II. VISUAL SEARCH ON WEB

World Wide Web, or Web search for visual (image, video) information is a part of modern life. A functional scheme for visual search of relevant images on Web is presented in Fig. 1.

Fig. 1. Functional scheme of visual search on Web.

In a Web visual search system, an image can be uploaded or selected as a sample. Such query image is further processed, where a search of similar and/or relevant images is performed. A user also makes interventions in the search system. Results of a search are presented to the user using Web technologies, for further processing and application.

A. Web Visual Search System

The Web Visual Search System proposed in this paper is presented in Fig. 2.

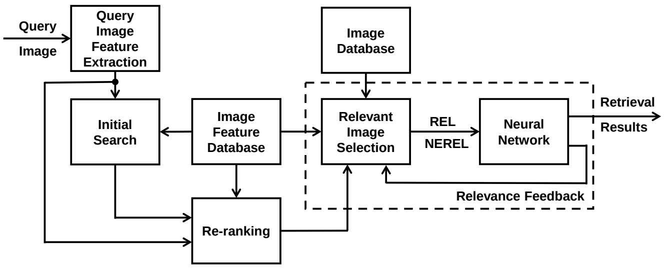

Fig. 2. The proposed Web Visual Search System.

An image that is used for the query in the Web Visual Search System can be uploaded or selected among the offered images. The query image is used as a referent sample in the system. The image is further processed in order to obtain adequate feature representation. Low-level image features are extracted within the system. Color, texture and shape description is used for the search procedure. If the image is selected from the dataset, features are already extracted.

Initial search in the system depends on the measurements of objective similarity between the feature vectors of the image query and other images from the database. Feature vectors for each image are obtained by concatenation of extracted features.

The result of the initial search is a set of images that are mostly similar to the query image based on the distance between the query feature vector and the feature vectors from the dataset. The Euclidean distance is used in this paper. The initial search result can be improved using re-ranking. In this step images are re-ranked on the basis of the query feature vector analysis and the analysis of the similarity between the query and other feature vectors according to their structure. The new set of images should be offered to the user as a choice to perform decision-making according to two bases: the global similarity based on the distance between the vectors and the local similarity of the vector structure. This set of images should enable the user to efficiently conduct the search in the desired direction. The next step in the search procedure is to perform relevance feedback and artificial intelligence mechanism in order to increase the efficiency of the search.

B. Machine learning and relevance feedback

For distance calculation between the feature vectors, Euclid distance is calculated. The distance describes the similarity between feature vectors. The comparison among the vectors is made in the objective and automatic manner.

After the comparison, the most L relevant images are presented to the user (L =50). This is followed by relevance feedback performed by the proposed system user.

User provides feedback in order to make further improvements in the search process. Labeling of the given set of L images is performed by the user. Each image is labeled as relevant (REL) or non (not)-relevant (NONREL), which represents the relevance feedback step.

The information about the relevance is further exploited by the system for further improvements in the search process. Intelligent logic implemented by machine learning is used for the correction of query feature vector. Artificial neural network is applied in order to take advantage of the REL-NONREL differentiation and to make the needed correction.

The next step is the initial search with the made correction over the query feature vector. As illustrated in Fig. 2 the procedure can be performed in several iterations until the user is satisfied with the obtained results.

Machine learning is implemented in the Web Visual Search system with relevance feedback using Radial Basis Function (RBF) neural network. Query feature vector displacement [7-8] and Gauss curve width change [9-10] are applied. Feature vector displacement is performed using query feature vector from the previous step, and the orientation and direction of feature vector displacement, based on the information gathered while differentiating images as REL and NONREL. The method's sensitivity can be modified by the change of Gauss curve width. The low variation clusters representing high similarity cases are intensified, and the queries are accordingly corrected. The machine learning approach represents here a sequence of feature vector corrections or modifications which iteratively displace a feature vector to an appropriate cluster according to similarity structure.

III. FEATURE ANALYSIS

Each image is represented using its color, texture and shape features. Feature vectors are obtained by the concatenation of extracted features. This set of features is normalized according to feature vector components using their maximum values. A brief description of features is given below, and methods for feature calculation is explained in detail in the literature [11].

A. Feature description

The color in images is described by color histogram descriptor. Hue-saturation value (HSV) color model is used with adequate quantization per channels (H:S:V=18:3:3), where totally 162 components are gathered for the purpose of feature description.

It is well known that color moments can be effective in representing color distributions in images. Thus, three color moments are used for image description: the first - mean, the second - variance and the third - skewness. For each of the channels in RGB image (R - red, G - green, B - blue) color moments are extracted. Also, images are divided into 9 non-overlapping blocks (3x3), obtaining 81 components for the description.

For the spatial correlation of colors, a color correlogram can be applied. It also gives an interpretation of the pixel color distributions. The correlogram is consisted of color pairs ($c1$, $c2$) for a pixel pair. The third coordinate is this model is the spatial distance between the pixels (d). Here, color autocorrelogram is applied, providing 648 components for the description.

In HSV color space, scalable color descriptor is used. It represents a color histogram encoded for efficient storage. Quantization is performed per each color channel. The ratio between the HSV coordinates is 16:4:4 (H:S:V). After scaling by one-dimensional Haar wavelet transform a set of 32 components is derived consisted of 16 lowpass coefficients and 16 highpass coefficients.

A color layout characterizes the spatial distribution of colors within image blocks. The blocks have size of 8x8 pixels. The color layout uses an array of representative colors for image blocks in *YCbCr* color space. Totally, 12 components are used for this color descriptor.

Radial co-occurrence is applied as a global descriptor for texture description. Four types of features like: entropy, energy, contrast and inverse difference moment for 24 different directions are extracted, obtaining 96 components for this description.

Wavelet transform is also applied for another texture descriptor. A wavelet texture grid descriptor is realized by dividing an image into 4x4 blocks. The fourth level of image decomposition is applied using Haar implemented as second generation wavelet. Totally, 192 components are obtained by gathering variance values of the high frequency sub-bands (12 sub-bands) for each block.

Besides texture, shape is also introduced as a relevant feature. Edge histogram and homogenous texture descriptors are applied. An edge histogram descriptor is obtained by dividing an image into 4x4 blocks. Five filters for five main directions are applied for each block as proposed in the MPEG-7 standard. This contributes the feature vector with 80 coordinates.

Homogenous Texture Descriptor is generated by tuned Gabor filters with 6 orientations and 5 scales. Directionality, coarseness, and regularity in the images are described using such approach, which contributes to the vector with additional 62 coordinates.

B. Re-ranking

The initial search step represents an objective measurement of the similarity among images from the database and the query image based on the Euclidean distance calculation.

Euclidean distance gives the possibility for introducing a variety of different features into the process of measuring the similarity among the feature vectors. Nevertheless, the dominance of certain features based on the employed length or dynamics of the intensity change among the components may mask the influence of other features. For this reason, there is a need to perform the appropriate correction of the initial search using the re-ranking, where images with more visual structural similarity with the query image are favored.

Re-ranking represents re-ordering the relevance of images after the initial search step based on the Euclidean distance as a measure of the similarity between the feature vectors. The image rating is determined using several parameters. The group of N = 100 images obtained by the initial search step is used for re-ranking. The first step in the ranking method is to determine the significance of the features within the feature vectors. For each feature standard deviation (STD) is calculated. Fig. 3 shows the main principles of the first step.

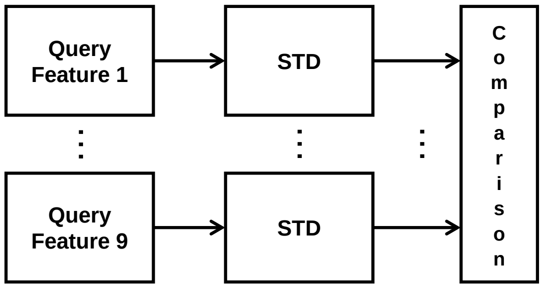

Fig. 3. The first step in re-ranking - the significance determination of query features based on the calculated feature standard deviation (STD).

The features in the query vector are compared using the standard deviation. The values of standard deviations FSD_i (i = 1, ..., 9) are sorted by their size, and for each feature an adequate rank FR_i is assigned. The value of the rank FR_i is the integer corresponding to the serial number of a feature in the sorted array FSD_i, starting from the lowest to the highest number (the highest number of FSD_i is 9). A greater significance is attributed to a feature with higher standard deviation. The assumption that a more dynamic change of feature components is more relevant for finding similar images in comparison to the features characterized by less variation is made.

The next step in the image ranking process is determining the significance of the features. The significance is based on percentage of the squared distance between particular features in the squared total distance between the query image and images from the group of similar images obtained by the initial search step.

Let K_i represent the percentage value. As a coefficient that describes the importance of the feature participation in differentiating the analyzed and the query image the value 1-

K_i is calculated. Thus, the image features with shorter distance in comparison to the query features are favored. So, the global similarity among the features is taken into account based on the Euclidean distance.

The third step in the re-ranking method is to compute the correlations between the image features from a group of similar images and corresponding features of the query image. The value R_i represents the correlation coefficient calculated for the i-th image feature from the group and the corresponding query image feature. Thus, the priority is given to the images with features characterized by greater structural similarity with the query image features.

The new rank from the group of similar images after the initial search is denoted as S_i, and it is calculated as:

$$S_i = FR_i(1-K_i)R_i,\ i = 1,...,9. \quad (1)$$

All the images from the group of similar images which are obtained after the initial search are sorted based on the new calculated rank S_i. The first M = 50 images are automatically selected after re-ranking, and forwarded to the user within the search system. The procedure includes user intervention and applying artificial intelligence in order to improve the search process.

IV. Experimental results

The search efficiency using re-ranking is analyzed here, and tested in the presented web visual search system. Two different scenarios are made.

The Scenario 1 represents image search, where after the initial search based on Euclidean distance, the obtained results are given to the user on the search correction using machine learning. The Scenario 2 represents a search where the initial search is corrected using the re-ranking method. After the re-ranking the obtained results are forwarded to the user for decision-making.

Dataset used for experimental analysis in this paper is Corel1000 [12] image database. This Corel image database is consisted of ten classes. The ten classes are differentiated according to its content and specific objects found in images. The names of the classes are: *Africa* (class1), *Beach* (class 2), *Buildings* (class 3), *Buses* (class 4), *Dinosaurs* (class 5), *Flowers* (class 6), *Elephants* (class 7), *Horses* (class 8), *Mountains* (class 9) and *Food* (class10). There are one hundred images per each class.

Search efficiency is tested for 1000 images found in the dataset. The efficiency is measured using the class affiliation. In each step during the search process, efficiency is defined as the ratio between the number of relevant images offered to the user for decision-making and the total number of images in the class. Efficiency is tested for five different search phases: the initial search and the four relevance feedback steps (search steps 1-5). Relevance feedback is performed automatically in the steps 2-5.

In Fig. 4 search efficiency is presented for each step in the search process. The results are averaged and presented for all 1000 images from the dataset.

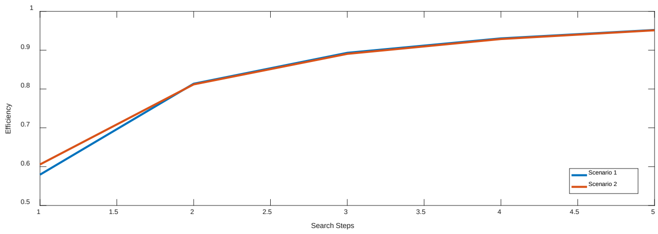

Fig. 4. Search efficiency measurement for Scenario 1 and Scenario 2, for all visual search steps: initial step and four relevance feedback steps. The results are averaged for 1000 images from the dataset.

In Fig. 4 it can be noticed that using the re-ranking method can generally increase the number of images belonging to the class of the query image in comparison to the Euclidean distance based initial search. During the relevance feedback process, the user takes control over the search process, so that both scenarios give similar performance results. The advantage of using the re-ranking is reflected in the structure of the images offered to the user for decision making. The most similar structures of the feature vectors are selected, so that the user can more effectively conduct further search procedure.

The results of the performance evaluation in the initial search are shown in Fig. 5. The results are averaged per each class.

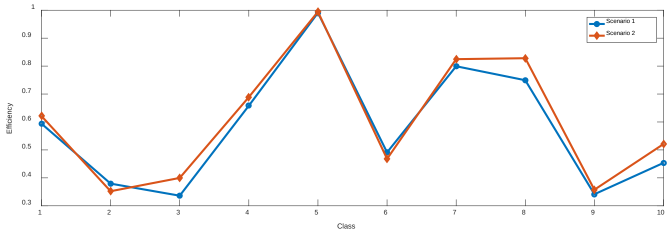

Fig. 5. Search efficiency in the initial phase, averaged per each class in the dataset for Scenario 1 and Scenario 2.

As shown in Fig. 5, the re-ranking method implementation generally enables greater search efficiency per class in the initial phase. The homogeneity of a class ensures the proximity of the feature vectors by the distance and the similarity of the feature vectors by the structure. So, the performance of the system in both scenarios are very similar. In the classes which contain images with similar objects, which are intuitively similar but significantly different in the feature vector structure, the re-ranking method has slightly lower efficiency than the Euclidean distance based method.

In Fig. 6 two images are presented from the class 2 - Beach. Both images intuitively belong to this class, but have significantly different feature vector structures.

Fig. 7 represents the measured search efficiency for the two images from the class 2 (Corel 104 and Corel 131). This class showed weaker performance for re-ranking in comparison to Euclidean distance method.

a) b)

Fig. 6. Examples of images from the class 2 in the dataset: a) Corel 104, and b) Corel 131 image.

Corel 104 image contains all the elements as most of the images in the class 2. The re-ranking gave the lower efficiency for Corel104 image in comparison to Scenario 1. The lower efficiency with the use of the re-ranking method in the initial phase does not represent an obstacle to exceed the scenario performance in the first step of the relevance feedback and using the Euclidean distance based method. The relevance of the offered images to the user greatly improves the search. In the case of the Corel 131 image, the re-ranking method generally enabled improved search efficiency within Scenario 2 in comparison to Scenario 1.

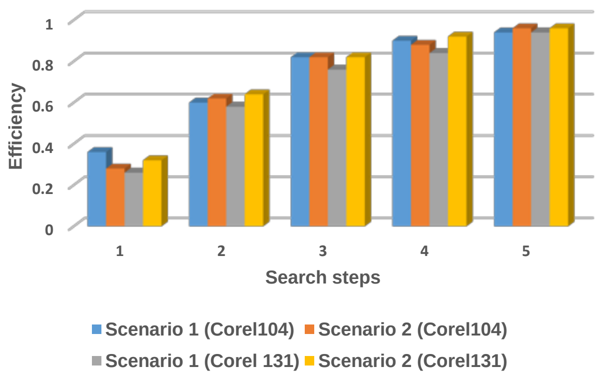

Fig. 7. Search efficiency in the case of two images (images Corel 104 and Corel 131).

By applying the re-ranking method, the search performance is generally improved in the image search system, which uses similarity among image feature vectors. The images that is given to user during the relevance feedback significantly affect the future search procedure. The re-ranking method makes relevant images. meaning images with similar feature vector structures, available to user in the web visual system.

V. Conclusion

In this paper, web visual search system is proposed with an improvement using re-ranking method. The system for image retrieval is based on an initial search using Euclidean distance. This is followed by further improvements using relevance feedback and artificial intelligence. By applying the re-ranking, the structure of images offered to a user for decision-making in relevance feedback process is improved. Namely, the images are ranked according to the feature vector corresponding to a query and its correlation with other images. The system is tested using Corel1000 dataset. The performance evaluation showed that re-ranking improves the efficiency of the initial search, which gives a user the ability to deal with a larger number of relevant images during visual information search.

Further work of the authors will be based on specific implementations, such as the proposed re-ranking, in order to achieve high performance in image clustering and obtain efficient search results.

References

[1] H. Hu, Y. Wang, L. Yang, P. Komlev, L. Huang, J. Huang, Y. Wu, M. Merchant, A. Sacheti, "Web-Scale Responsive Visual Search at Bing," *arXiv preprint arXiv:1802.04914*, 2018.

[2] F. Yang, A. Kale, Y. Bubnov, L. Stein, Q. Wang, H. Kiapour, and R. Piramuthu, "Visual search at ebay," In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 2101-2110, August 2017.

[3] G. Zajić, N. Kojić, V. Radosavljević, M. Rudinac, S. Rudinac, N. Reljin, I. Reljin, B. Reljin, "Accelerating of Image Retrieval in CBIR System with Relevance Feedback", EURASIP Journal on Advances in Signal Processing, Spec. Issue on Knowledge Assisted Media Analysis for Interactive Multimedia Applications, Volume 2007 (2007), Article ID 62678, 13 pages(doi:10.1155/2007/62678).

[4] W.H. Hsu, L.S. Kennedy and S.F. Chang, "Reranking methods for visual search," *IEEE MultiMedia*, *14*(3), 2007.

[5] X. Yang, T. Mei, Y. Zhang, J. Liu and S.I. Satoh, "Web image search re-ranking with click-based similarity and typicality," *IEEE transactions on image processing*, *25*(10), pp.4617-4630, 2016.

[6] G. J. Zajić, N. S. Kojić, and B. D. Reljin, "Improving CBIR Systems Using Automated Ranking," *Telfor Journal*, Vol. 4, No. 2, 2012, pp.116-121.

[7] Y. Rui, T. Huang and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS", in *Proc. IEEE Int. Conf. on Image Processing (ICIP '97)*, vol. 2, pp. 815–818, Santa Barbara, CA, USA, October 1997.

[8] [12] P. Muneesawang, and L. Guan, "An interactive approach for CBIR using a network of radial basis functions", *IEEE Trans.Multimedia*, vol. 6, pp. 703-716, October 2004.

[9] Y. Rui and T. Huang, "Relevance feedback: A power tool for interactive content-based image retrieval", *IEEE Trans. Circ. Syst. for Video Technology*, vol. 8, pp 644–655, 1998.

[10] J. Peng, B. Bhanu, and S. Qing, "Probabilistic feature relevance learning for content-based image retrieval", *Computer Vision and Image Understanding*, vol. 75, pp. 150–164, 1999.

[11] N. Kojić, G. Zajić, S. Čabarkapa, M. Pavlović, V. Radosavljević, and B. Reljin,"The Influence of the Feature Vector Content on the CBIR System Efficiency," TELSIKS 2009, *9th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services*, Nis-Serbia, 7.-9. October 2009.

[12] Corel 1000 database, http://wang.ist.psu.edu/docs/related/, last accessed 15.03.2018.