Re-ranking of Image Search Results using a Graph Algorithm

Hilal Zitouni, Sare Sevil, Derya Özkan, Pınar Duygulu
Bilkent University, Department of Computer Engineering
{zitouni, sareg, duygulu}@cs.bilkent.edu.tr, derya@usc.edu

Abstract

We propose a method to improve the results of image search engines on the Internet to satisfy the users who desire to see the relevant images in the first few pages. The results of the text based systems, that use only the accompanied text of the images, are re-ranked by incorporating the visual similarity of the resulting images.

We observe that, in general, together with many unrelated ones, the result of text based systems include a subset of correct images, and this set is the largest most similar one compared to other possible subsets. Based on this observation, we present the similarities of all the images in a graph structure, and find the largest densest component of the graph, corresponding to the largest set of most similar subset of images. Then, to re-rank the results, we give higher priority to the images in the densest component, and rank the others based on their similarities to the images in the densest component. The experiments carried out on 10 category of images from [4] promise the success of our method over Google ranking.

1. Introduction

Image search engines on the web do not use visual content of images and therefore perform poorly. The text based results are likely to provide many unrelated images together with a small subset of relevant images since the accompanying text used in searching may be irrelevant or errorful. It is also very common to have synonyms of a word causing multiple visual categories to be mixed in the resulting set. On the other hand, users want to see visually similar images corresponding to their query in the initial pages of the search results. However, performing a fully image based search engine by recognizing objects and scenes still is not in the capability of computer vision systems.

In this study, we propose a method to satisfy the users of image search engines by re-ranking the results of text based systems using visual information. In our approach, with the assumption that there will be a large set of visually similar images relevant to the query in the set of all resulting images, we find the largest set of most similar images and place them in the earlier pages.

There are some other studies attacking a similar problem [2, 3, 4, 5]. The method, ReSPEC, proposed in [3], takes a subset of images and segments all images in to blobs. When clustered, densities of blob clusters become directly proportional to the relevancy of images in that cluster. Using this idea, remaining images are inserted to appropriate clusters and images are re-ranked. The objective of the work by Schroff et. al. [4] is to form categorized image databases 'harvested' from the web. The re-ranking operation is performed for separating relevant and irrelevant results by the usage of a combination of textual and visual features.

Generalizing the method that we previously used to find the relevant faces associated with a name [6], in this study, we propose a graph based approach to find the group of relevant images in the result of a text based search. In our approach, rather than multiple clusters, we aim to have a single cluster corresponding to the relevant set even in the existence of large variations in the category.

In our approach, similarity of images, computed by using the interest points extracted over the images, are represented in a fully connected graph structure Then, the problem of finding the most similar group of images turns into the problem of finding the largest densest component in the graph. For this purpose we

Preprint submitted to 19th International Conference on Pattern Recognition.
Received April 16, 2008.

utilize the greedy densest component algorithm proposed by Charikar []. The images located in the densest component are assumed to be the relevant images, and given higher priority and the rest of the images are ranked according to their similarity to the images in the densest component.

In the following we will describe the details of the algorithm. We should note here that, the proposed algorithm only make use of the available data, and does not require a supervised input to specify the relevant images.

4. Approach

Our method consists of the following steps: First, interest points on the images are extracted, and the similarity of each pair of images are computed based on the similarity of matching interest points. Then, in order to reduce the time complexity of the algorithm and increase the reliability, rather than the entire result set of images, a subset from the first few pages of the original results is taken as a model for constructing the graph. In the fully connected graph constructed over this subset, the nodes are the images in the model, and the edges are the similarity of images. Note that, the selection of the model is arbitrary and unsupervised. The only important factor for the proposed algorithm is to have a sufficiently large number of images corresponding to the most similar set among the set of images selected for building the initial graph. Therefore, to build the initial model, in the experiments, we choose the first 30 images with the assumption that the first pages will still include the most relevant images despite the errors of text based search results. In the next step, the original graph with real valued edges is converted into a binary graph in order to apply the greedy densest component algorithm of Charikar [1]. Each time by removing a single node from the graph, the algorithm decides on the largest densest component of the graph. In the final step, the images in the densest component are placed into the higher ranks, and the rest of images that was left previously out of the model are ordered according to their similarities to the densest component. In the following each step will be described in more detail. Figure 1 shows the overall method.

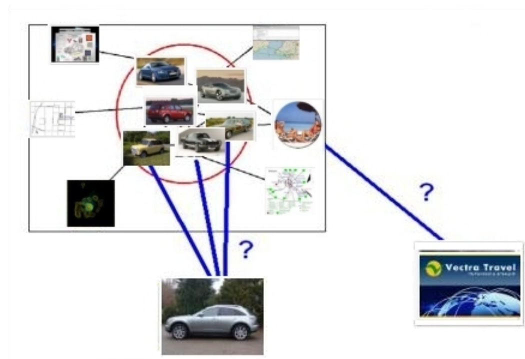


Figure 1. Overall algorithm.

Construction of the similarity graph :
We represent the similarity of images using the matching interest points extracted over the images. We make use of SIFT operator proposed by Lowe [7] to detect and describe the interest points. However, rather than using the original matching criteria which results in small number of matches for the images on the web having large variations, we propose a new matching scheme. For each point in one of the image pairs, the Euclidean distance is used to find the best math from the other image with the minimum distance. This approach assigns a match to all of the points in the images. In order to eliminate the wrong matches, we apply a uniquness constraint [6] which satisfies that there will be a unique, one-to-one match between the pairs of points, and the others not satisfying this condition will be eliminated. The similarity of two images is then computed as the average distance of the matching interest points. Finally, a fully connected similarity graph is constructed with the images being the nodes, and the edges being the similarity of images.

Forming the binary graph suitable for Charikar's Algorithm:
Charikar's greedy densest component algorithm [1] requires a binary graph. The real-valued graph is converted into binary form by selecting a threshold, and converting the edges above threshold to 1, and the others to 0. In the experiments the effet of threshold is tested and, found that although not affecting the results in a serious way, the choice of 0.3 is a satisfactory one. Finally, the 1 edges are kept, and the others are removed to feed the graph into the densest component algorithm.

Preprint submitted to 19th International Conference on Pattern Recognition.
Received April 16, 2008.

Finding the Densest Component:
This newly formed graph is given to Charikar's algorithm and the densest component is found for that graph. Charikar's algorithm can be summarized as follows: a density value is computed for all subsets of the given graph obtained by removing one node each time. Then the subset with the largest density is selected as the densest component. For a given subset S of the similarity graph, the density is computed as follows:

$$f(S) = \frac{|E[S]|}{|S|},$$

where $f(S)$ is the density of S; $|E[S]|$ is the number of edges in S; and $|S|$ is the number of nodes in S.

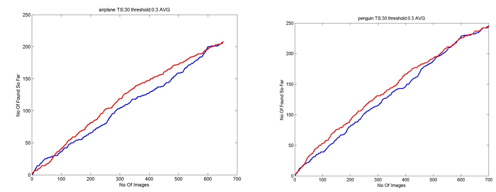
Expanding the model with the other images:
The densest component is found as described above on the subset of images which we refer as the model. Then, the remaining images are ranked according to their similarity to the elements of the densest component.

5. Experiments

In order to measure our performance we have used the dataset provided by Schroff et al. [4] which consists of images harvested from Google's Image Search results.

In order to form the fully connected graph and to extract the initial densest component from the graph, we choose a subset from the Google Image Search results. In the experiments, we have found that 30 images as a model size is generally sufficient to capture the densest component. Then, as it was explained in the previous section, remaining images in Google Image Search results are re-ranked.

In order to visualize our performance we have plotted recall versus precision graphs for some categories in Figure2. Because of the difference of our method to the method presented in [4], we were not able to compare our performance to that performance. We were, however, able to compare our performance to Google's Image search. From the below graphs one can see that our performance has been, mostly, slightly better than Google's image search.



(a) airplane *(b) penguin*

Figure 2. Recall vs. *precision graphs some categories. Red line: performance of our method; Blue line: performance of Google's image search algorithm.*

Table 1 shows the Mean Average Precision values for our method compared to the ones of Google's performance for the 10 different categories. Figures 3 and 4 show the first 20 images for airplane ategory obtained from Google rank and re-ranked with our approach.

Table1: mAP values for comparison on 10 categories

Category	Google results	Our results
Airplane	0.3555	0.3809
Car	0.4365	0.4661
Penguin	0.4002	0.4397
Wristwatch	0.7777	0.7709
Camel	0.3786	0.3710
Boat	0.4078	0.3894
Guitar	0.5312	0.5437
Elephant	0.4062	0.3982
Motorbikes	0.5700	0.5445
Bikes	0.4633	0.4271

6. Summary and discussion

In this study, we propose a method to re-rank the Google's image search results. The proposed method make use of the observation that, although the original results include many irrelevant images, still the largest most similar subset of these images should correspond to the query. With the representation of image similarities in a graph structure, we converted the problem into the finding of densest component in the graph.

Preprint submitted to 19th International Conference on Pattern Recognition.
Received April 16, 2008.

The results are promising, with some categories producing better results than the Google search especially for the first few pages where the users want to see the most relevant images.

The worse results obtained for some categories may due to two reasons. First of all, the proposed method assumes that in the model selected for finding the densest component, there are more instances of the query image compared to the others. When this is not the case, the proposed method may result in a wrong densest component. The second reason may be the features used the the experiments. For some categories, especially with smooth surfaces, the interest point based matching does not provide a good similarity measure. In the future, we plan to use other features including color and texture to improve the results.

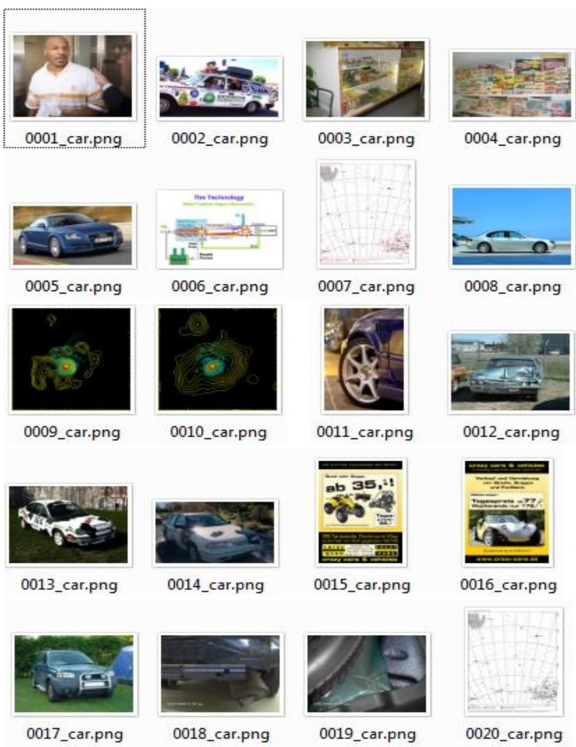


Figure 3. Google's ranking

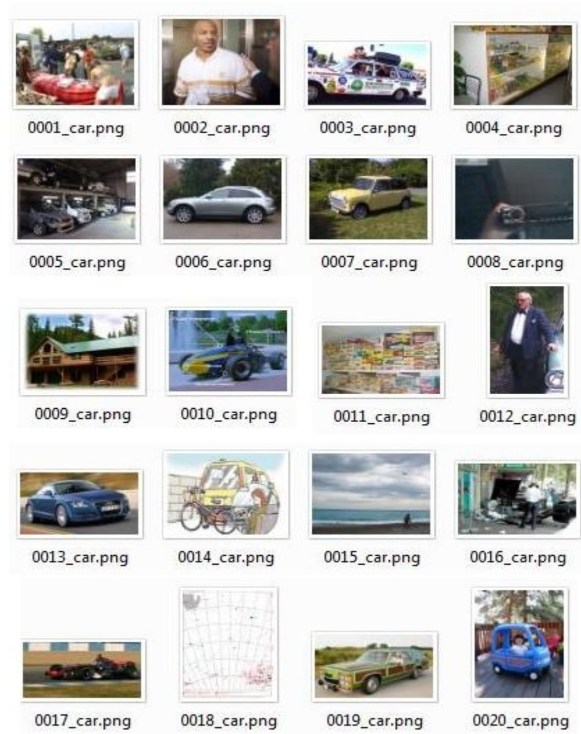


Figure 4. Ranking of our method.

References

[1] Moses Charikar. Greedy approximarion algorithms for finding dense components in a graph. In *APPROX' 00: Proc. Of the 3rd International Workshop on Approximation Algorithms for Combinatorial Optimization*, London, UK, 2000.

[2] Winston H. Hsu, Lyndon S. Kennedy, Shih-Fu Chang. Video search reranking via Information Bottleneck Principle.

[3] Nadav Ben-Haim, Boris Babenko and Serge Belongie. Improving Web Based Image Search via Content Based Clustering. [IEEE Xplore] *SLAM 2006,* New York City

[4] F. Schroff, A. Criminisi, and A. Zisserman. Harvesting Image Databases from the Web. *ICCV 2007*

[5] T. L. Berg and D. A. Forsyth. Animals on the Web. In *Proc. CVPR,* 2006.

[6] Derya Özkan, Pınar Duygulu. A Graph Based Approach for Naming Faces in News Photos. In *Proc. CVPR,* 2006.

[7] D. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2). 2004

Preprint submitted to 19th International Conference on Pattern Recognition.
Received April 16, 2008.