UAIC's participation at Wikipedia Retrieval @ ImageCLEF 2011

Emanuela Boroş, Alexandru-Lucian Gînscă, Adrian Iftene

UAIC: Faculty of Computer Science, "Alexandru Ioan Cuza" University, Romania
{emanuela.boros, lucian.ginsca, adiftene}@info.uaic.ro

Abstract. This paper describes the participation of UAIC team at the ImageCLEF 2011 competition, Wikipedia Retrieval task. The aim of the task was to investigate retrieval approaches in the context of a large and heterogeneous collection of images and their noisy text annotations. We submitted a total of six runs, focusing our effort along the textual retrieval, query expansion on English language, combined with feature extraction (Color and Edge Directionality Descriptor, CEDD). Our intention was to build a CBIR (Content-based image retrieval) system that relies on a fast indexing and retrieval practice based not only on the textual multilingual metadata, but also on the images features. The results were satisfying in the multilingual mixed search (text and images) and query expansion approach.

Keywords: Support Vector Machine, Query Expansion, Similarity Identification, Content-based image retrieval

1 Introduction

It is a truism to observe that images are currently used in all types of applications. Methods are needed to be used to summarize, describe and classify collections of data. The influence of visual perspective in today's society is clear for all to see. The difficulty of locating a desired image in a large data collection increased the need for a CBIR (Content-based image retrieval) with more effective techniques.

The subject of this paper outlines our approach for the Wikipedia Retrieval[1] task at ImageCLEF 2011[2]. The task addresses the investigation on retrieval approaches in the context of a large and heterogeneous collection of images and their noisy text annotations (similar to those encountered on the Web) that are searched for by users with diverse information needs. We received a collection of images that consisted of 237,434 images and associated user-supplied annotations, built to cover similar topics in English, German and French and a file in XML format with the topics and their metadata. This study was performed in order to build a system capable of classifying, indexing and retrieving images from a large-scale collection of images.

The paper is organized as follows: in Section 2 we describe our system, while the advantages, results of the system and experimental results are reported in Sections 3

[1] Wikipedia Retrieval task: http://imageclef.org/2011/wikipedia
[2] ImageCLEF 2011: http://imageclef.org/2011

and 4. Last Section draws conclusions regarding our participation in Wikipedia Retrieval task at ImageCLEF 2011.

2 System Description

One of the main problems of building such a system is the difficulty of locating a desired image in a large data collection. While it is perfectly feasible to identify a desired image from a small collection simply by browsing, more effective techniques are needed with collections containing thousands of items. Access to a desired image from a large collection might thus involve a search for images depicting specific types of features, evoking a particular mood, or simply containing a specific texture or pattern. Query by example is a query technique that involves providing the system with an example image that it will then base its search upon. The underlying search algorithms may vary depending on the application, but result images should all share common elements with the provided example.

Our system can be divided into two main components, one for text processing and one for image processing. All components were written in Java with the use of Lucene[3] library and LIBSVM[4], a library for support vector machines [1]. The choice of the classifier is a key ingredient for an effective machine learning based image recognition system. We chose Support Vector Machines (SVMs) [2] based on their state-of-the-art performances in several visual recognition domains and Lucene for its utility in the implementation of Internet search engines and local, single-site searching. For image analysis, we used CEDD features [3] that were donated by the DUTH team (i.e., the Information Retrieval Unit, Department of Electrical and Computer Engineering, Democritus University of Thrace, Greece) for this year's competition and for text analysis we used the full-featured text search engine library Lucene. A main interest was to create a gold trunk of images so a more reliable classification could be done.

The architecture of the system relies on a model-view-controller pattern, the controller being multithreaded and can accomplish more requests at a time. It manages the feature extracting module and the final result computing module. Therefore, one of the maturity levels that the system has is the possibility of retrieving images in real-time. The main focuses in this project are data representation, feature extraction and indexing, image query matching and user interfacing. We can consider a training stage with the sample image collection. The main flow of the application is described in Figure 1.

[3] Lucene: http://lucene.apache.org/java/docs/index.html
[4] LIBSVM: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

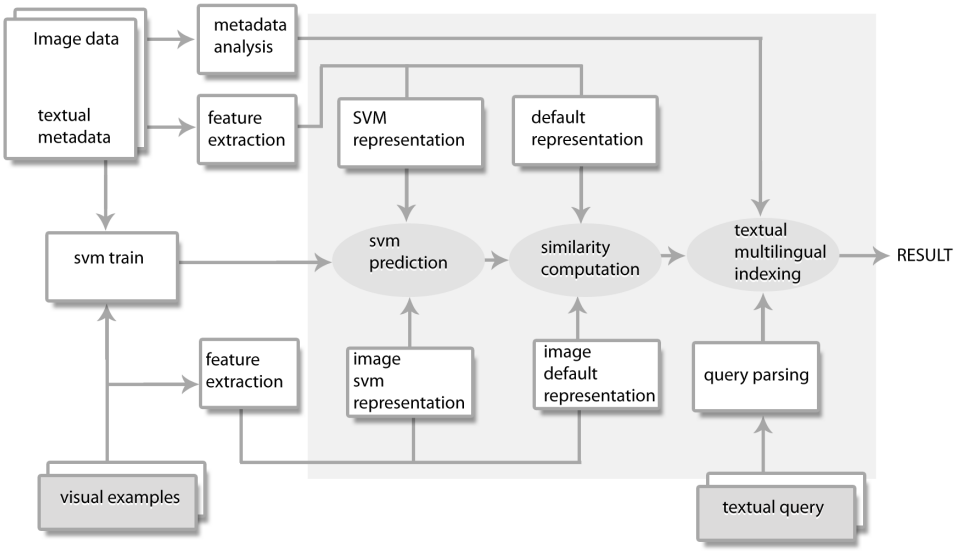

Figure 1: Basic system used for image retrieval with query-by-example**.**

2.1 Image processing

A model construction step aims at providing a representation of the input data that minimizes the within-class variability while at the same time maximizing the between-class variability in this type of system, CBIR (Content-Based Image Retrieval) [4]. Additionally, this representation is usually more compact than raw input data and therefore allows reducing the computational load imposed by the classification process. When the user can provide example images which contain instances of the visual properties, content or configurations they would like to search for, it is very hard for the system to ascertain which aspects make a given image relevant and how similarity should be assessed. Many such systems therefore rely on extracting different images features to guide the search towards desirable images, but this approach is still studied so it would work in real-world retrieval scenarios.

Features can be derived from the whole image (global features) or can be computed locally, based on its salient parts (local features). We chose one scale-space theory based features, a global one (Color and Edge Directionality Descriptor, CEDD) and no local one considering that the most influential factor of the query result will be the textual metadata. The rest of the section describes briefly this approach. The SIFT (Scale Invariant Feature Transform) components [5, 6] were used similar with approach from [7].

First of all, the CEDD features for the large data collection are already extracted thanks to Information Retrieval Unit, Department of Electrical and Computer Engineering, Democritus University of Thrace, Greece and we also extract the CEDD features from example images. Every image is represented by a file of approximately 54 bytes, meaning a vector of 144 numbers in the default representation. We modify this representation to a specific one needed by LIBSVM. Search results are sorted based on their distance to the queried images. It is virtually impossible to compare

images using traditional methods such as a direct comparison between gray values so we chose the more practical method for comparison, a simple similarity function. An image distance measure compares the similarity of two images in various dimensions such as color, texture, shape, and others. So, having the CEDD representation of the images collection and for the example images, the most common method for comparing two images (an example image and an image from the large collection) is Euclidean distance similarity function.

For the second representation, the SVM trainer is learning the example images and then it predicts the category for every image in the large data collection. The given topics contained an XML with fifty topics with their metadata and example images. From the example images, gold images collection is constructed and used for basic comparison with CEDD features and for training data for SVM classifier, as it can be seen from Figure 1. This is considered the training step, so, for the SVM classifier, we obtained fifty classes.

2.2 Text processing

In a first step, we extracted the textual information from the metadata files and combined the data from the description, comment and caption tags. This was done independently for each of the three languages. We then created a Lucene index for each language. A separate index was created for English, in which the text was put through a stemming process, using the Porter Stemmer[5] provided by the Lucene framework[6] and a stop word elimination process.

The topics are used as Lucene search queries after they go through a processing step. For English, we generated a separate query that consists of the stemmed initial text. Also, we give a special attention to capitalized words. We consider a series of capitalized words to be a named entity and we give it an increased boost in the Lucene query. Due to the nature of the short, simple topics, we didn't find the need to use a gazetteer based named entity recognition approach. For some runs, we have used a simple form of query expansion. This consisted of introducing synonyms with the help of WordNet[7] and using them as a disjunction of terms consistent with the Lucene query syntax.

2.3 Aggregation of results

There are two places in our system where intermediary results are combined. The first addressees the multilingual runs and gives a method for selecting the top answers for the queries posed in each of the three languages, while the second bridges the results obtained after the textual queries with those given by the image matching side of the system.

[5] Porter Stemmer: http://tartarus.org/~martin/PorterStemmer/
[6] Lucene: http://lucene.apache.org/java/docs/index.html
[7] WordNet: http://wordnet.princeton.edu/

Using queries in all of three languages, English, French and German, imposed the problem of retaining the best results for each language. This was done by first obtaining a maximum of 1000 documents sorted descending by the Lucene score for each language. In the next step, all of the documents were mixed and reordered by their associated scores and after duplicate documents were eliminated, the first maximum 1000 documents were returned. This process is sketched in Figure 2.

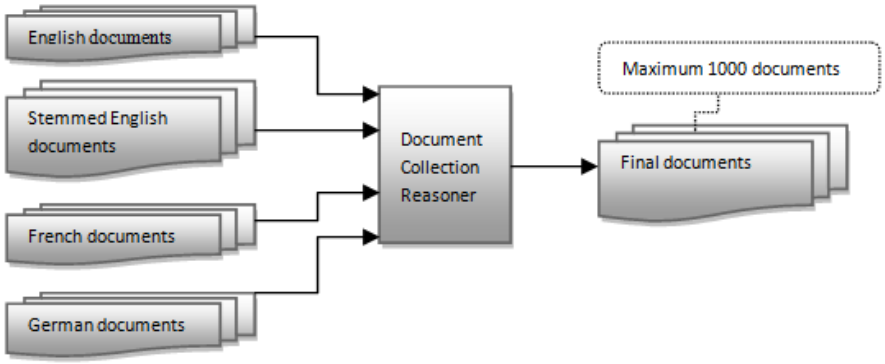

Figure 2: Textual query results aggregation

The second time when aggregation of partial results is required is after the final documents from the textual queries are obtained and the similarity scores for the topic images are calculated. To obtain a uniform set of final scores, the similarity distance score are brought in the [0, 1] interval. After multiple tests, we settled on a weighted mean between the Lucene scores and the normalized image similarity scores, with the first having a contribution of 60% of the final score.

3 Experiments

This section describes the setup used for the experiments reported in this paper. We conducted six experiments to evaluate the effectiveness of our approach. In all of them, we compared the techniques as well as the batch algorithm. We employed Lucene library and LIBSVM library, and we chose multiclass SVM. The local descriptors that we already discussed, separated in files with the specific SVM representation, are used as input to SVM via the RBF kernel and the training data that contained fifty classes described by CEDD descriptors extracted in model construction phase. First results are obtained from textual indexing; the second results from similarity computation between example images and the large data collection and the third results are obtained from SVM prediction.

In all the experiments, we benchmarked against a system not using any prior knowledge. The prior-knowledge model was built from the example images. The six experiments are separated by analysis on different languages, SVM use and image processing. Three runs were submitted for English only queries and the other three for

all languages. The multilingual scores of images were formed by adding up scores in individual languages. As with multimodal runs, the overall performance was increased with respect to English only runs. Every experiment had different results from different sources (textual or visual) and the overall classification rate was then computed as an average, to which the results from each method contributed equally. Every experiment had different results from different sources (textual or visual) and score between 0 and 1 is attributed to the image. The overall classification rate was then computed as an average, to which the results from each method contributed equally.

4 Results

All the obtained results are summarized in Table 1. All submitted runs were automatic and combined, so an initial text search is followed by a possibly combined visual search. We could have obtained better results if we had combined another SVM parameters, the local descriptor (SIFT) and query expansion on all languages. The presented results provide clear evidence of the capability of the query expansion method and textual retrieval on all languages to perform a satisfying search result.

Table 1: UAIC runs in Wikipedia Retrieval 2011 task[8]

	group	runid	mode	FB/QE	Annotation language	Topic language	MAP	P10	P20	Rprec	bpref
94	UAIC2011	uaic3lucene&alllang&cedd	Mixed	NOFB	EN+FR+DE	EN+FR+DE	0.1665	0.4080	0.3090	0.2313	0.1934
101	UAIC2011	uaic4lucene&svm&alllang&cedd	Mixed	NOFB	EN+FR+DE	EN+FR+DE	0.1259	0.3340	0.2430	0.1817	0.1516
103	UAIC2011	uaic6lucene&alllang&qe&cedd	Mixed	QE	EN+FR+DE	EN+FR+DE	0.1099	0.2840	0.2250	0.1552	0.1305
104	UAIC2011	uaic4lucene&svm&en&cedd	Mixed	NOFB	EN	EN	0.0955	0.2800	0.2100	0.1587	0.1236
105	UAIC2011	uaic2lucene&cedd	Mixed		EN		0.0946	0.2800	0.2110	0.1629	0.1249
106	UAIC2011	uaic5lucene&en&qe&cedd	Mixed	QE	EN	EN	0.0705	0.2380	0.1820	0.1158	0.0936

In this section, we analyze strengths and weaknesses of the system. For that we compare the results obtained for two of our better runs: *uaic3lucene&alllang&cedd* and *uaic6lucene&alllang&qe&cedd*. The first experiment contributed to the best result by using metadata in all languages indexed by Lucene. In the second experiment we only added query expansion on English language. The experiment where we used SVM prediction had not satisfying results. So, the success rate for the ones with textual retrieval on all languages is better than in the other cases where SVM or only English language metadata was used. Document expansion can improve the MAP from 0.1099 to 0.1665, but query expansion in combination with our methods does not show much improvement.

[8] ImageCLEF 2011: Wikipedia image retrieval results: http://www.imageclef.org/2011/wikimm-results

5 Conclusions

In this paper we describe our first participation at Wikipedia Retrieval at ImageCLEF 2011 competition. We have experimented with different methods combined. However, the obtained results show that it is necessary to continue investigating the expansion methodology, to better apply any content-based image retrieval techniques that helps us to extract the features of the images, to improve our SVM learner. Thus, our next goal will be to improve the expansion by applying some more techniques. For example, it will be interesting to apply query expansion on all languages and repair our computation function for combining all methods. Our runs got bad results due to some computation error which will fixed in the future research. In addition, further investigation in textual processing could achieve better results.

Acknowledgements. The research presented in this paper was funded by the Sector Operational Program for Human Resources Development through the project "Development of the innovation capacity and increasing of the research impact through post-doctoral programs" POSDRU/89/1.5/S/49944.

References

1. Chang, C. C., Lin, C. J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27. (2011)
2. Hsu, C. W., Chang, C. C., Lin, C. J.: A Practical Guide to Support Vector Classification. At Department of Computer Science, National Taiwan University, Taipei 106, Taiwan, April 15. (2010)
3. Chatzichristofis, S., Boutalis, Y. S.: CEDD: Color and Edge Directivity Descriptor – A compact descriptor for image indexing and retrieval. In 6th International Conference in advanced research on Computer Vision Systems ICVS 2008, May 12-15, 2008, Santorini, Greece. (2008)
4. Eakins, J., Graham, M.: Content-based Image Retrieval. Report number 39 at University of Northumbria at Newcastle, October. (1999)
5. Lowe, D. G.: Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision. Corfu, Greece. Pp. 1150-1157. (1999)
6. Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision. 60, 2. Pp. 91-110. (2004)
7. Boroş, E., Roşca, G., Iftene, A.: Using SIFT Method for Global Topological Localization for Indoor Environments. In C. Peters et al. (Eds.): CLEF 2009, LNCS 6242, Part II (Multilingual Information Access Evaluation Vol. II Multimedia Experiments). Pp. 277-282. ISBN: 978-3-642-15750-9. Springer, Heidelberg. (2009)