Some Results Using Different Approaches to Merge Visual and Text-Based Features in CLEF'08 Photo Collection

Ana García-Serrano[1], Xaro Benavent[2], Ruben Granados[3], and José Miguel Goñi-Menoyo[3]

[1] Universidad Nacional de Educación a Distancia, UNED
[2] Universidad de Valencia
[3] Universidad Politécnica de Madrid
agarcia@lsi.uned.es, xaro.benavent@uv.es, rgranados@fi.upm.es, josemiguel.goni@upm.es

Abstract. This paper describes the participation of the MIRACLE team[1] at the ImageCLEF Photographic Retrieval task of CLEF 2008. We succeeded in submitting 41 runs. Obtained results from text-based retrieval are better than content-based as previous experiments in the MIRACLE team campaigns [5, 6] using different software. Our main aim was to experiment with several merging approaches to fuse text-based retrieval and content-based retrieval results, and it happened that we improve the text-based baseline when applying one of the three merging algorithms, although visual results are lower than textual ones.

Keywords: Information Retrieval, Text-based image Retrieval and Content-based image Retrieval, Visual features, Textual features, Merge result lists, Indexing.

1 Introduction

MIRACLE is a consortium formed by research groups from different universities in Madrid, Universidad Politécnica (UPM), Universidad Autónoma and Universidad Carlos III, along with DAEDALUS, a SME spin-off of UPM. This paper describes our participation at the ImageCLEF Photographic Retrieval task of CLEF 2008, fully described in [1, 2]. This campaign Mir-FI team (MIRACLE at UPM) joined the Vision-Team at the University of Valencia (UV) who has developed a Content-Based retrieval system (CBIR) [4], in which the low-level features have been adapted to be used at the ImageCLEFphoto.

We succeeded in submitting 41 runs with results obtained by using (1) our re-implemented module for textual retrieval based on the classical vector model (VSM) in Information Retrieval, (2) the content-based image module (developed by UV)

[1] This work has been supported by the Spanish R+D National Plan, projects TIN2007-67407-C03-03, TIN2006-10134 and by Madrid R+D Regional Plan, project MAVIR, S-0505/TIC/0 00267 (www. mavir.net).

C. Peters et al. (Eds.): CLEF 2008, LNCS 5706, pp. 568–571, 2009.
© Springer-Verlag Berlin Heidelberg 2009

with five different methods for aggregation, and (3) the three new merging algorithms using textual and visual features.

Obtained results from text-based retrieval are better than content-based ones. By merging both textual and visual retrieval we improve the text-based one when applying one of the merging algorithms implemented (the so-called ENRICH).

2 Detailed Description of Experiments

This year Mir-FI system allows executing the different configurations that are explained in the following.

MIRACLE-FI textual retrieval is based on the VSM approach using weighted vectors based on the TF-IDF weight. The implemented components are: (1) The Text Extractor, (2) the Preprocessor (to special characters deletion and Stop-word detection), the (3) Annotations/Topics Tags Selector, to select tags from the annotations files (TITLE, DESCRIPTION, NOTES and LOCATION), and from the topics (TITLE and NARR), (3 and 4) MirFi-VSM Indexer and MirFi-VSM Searcher (applied on the associated text to the images and using a slightly different cosine measure).

The CBIR use different low-level features describing (a) Color information with a histogram of the HS (Hue, Saturation) values of the image pixels (quantization of the HSV space into 30 color bins) and also describing (b) Texture information using different feature textures (Gabor Convolution Energies, Gray Level Coocurrence Matrix also known as Spatial Gray Level Dependence, Gaussian Random Markov Fields, the granulometric distribution function and the spatial distribution).

Secondly the CBIR module calculates the similarity distance between the feature vectors from each image on the database to the three topic images. The two distance metrics used are: the Euclidean and the Mahalanobis. The so-called OWA [4] operators have been used to aggregate the three low-level feature vectors of the topic images.

Finally, textual and image results lists are merged in three different ways:

FILTER-N. This way of merging the image and textual results lists consists on checking which results in the textual results list are also included in between the N first results of the visual results list. The value of N indicates the number of results taken into account from the visual list when narrowing down the textual one. The resulting merged list will have a maximum of 1000 results for each query.

ENRICH. Also uses two results lists, the main and the support list. If a concrete result appears in both lists for the same query, the relevance of this result in the merged list will be increased in the following way:

$$new\,\mathrm{Re}\,l = main\,\mathrm{Re}\,l + \frac{\sup \mathrm{Re}\,l}{\left(pos\,\mathrm{Re}\,l + 1\right)} \tag{1}$$

where newRel is the relevance value in the merged list, mainRel is the relevance value in the main list, supRel in the support list, posRel is the position in the support list. Relevance values will be then normalized from 0 to 1.

Every results appearing in the support list but not in the main one (for each query), will be added at the end of the results for each query (a maximum of 1000 results per

query). In this case, relevance values will be normalized according with the lower value in this moment.

TEXT-FILTER. In this case, the textual module is applied to the complete database and only those images that have a relevance value above zero are passed to the CBIR. Then, the CBIR calculates the distance similarity to each of the three query and these values are merged with the different OWA aggregation operators.

Table 1. Our best results for each of the different runs we participated

Run Identifier	Textual Retrieval	Visual Retrieval			P20	MAP
		Distance	Merge topics	Merge		
TXT-baseline	SVM	--	--	--	0.2253	0.2846
IMG-mahamin	--	maha	min	--	0.0213	0.0679
TXTIMG-merge06mahamin	SVM	maha	min	ENRICH	0.3090	0.2401
TXTIMG-criba10000mahamin	SVM	maha	o3	FILTER-10000	0.3179	0.1936
TXTIMG-merge06mahamin	SVM	maha	min	ENRICH	0.2401	0.3090

3 Runs and Results

Finally, it was submitted one text-based run, 10 content-based runs and 30 mixed runs using a combination of both. The details can be found in [2].

In the general classification with all the automatic runs (1039) from all the participant groups (25), our best position was obtained is 306, corresponding to the English automatic textual and visual retrieval run using the ENRICH merge, being our best values not far away from the bests values from all the automatic experiments; we were even better if taking the best 4 runs from each participating group. The English automatic textual retrieval module with no linguistic processes, that is our "baseline" run, appears in the position 185 (over 399). The content-based image module results show that Mahalanobis distance outperforms the Euclidean one, and the best aggregation method in both metrics is the minimum (AND), followed by the orness(W)_0.3 that is a smoothed AND. Our best result for this group of experiments was the combination of Mahalanobis metric with orness(W)_0.3 aggregation method, and was considerably lower than the best results. The merge results in which we experimented with the 3 different algorithms, are: FILTER-10000 that improved the text-based baseline in low precision values (P5, P10 and P20), but never MAP neither number of relevant images retrieved. The run English automatic textual and visual retrieval with Mahalanobis metric and the min operator, obtained the best P20 value from all ours experiments (190th in the general P20 classification, over 1039).

Experiments applying ENRICH improved the baseline in MAP and in number of relevant images retrieved. Our best MAP was achieved merging the textual results with the visuals ones obtained using the Mahalanobis metric and the AND operator which is in the 91st position in the general MAP classification (over 1039). This value is higher than the average MAP taken from the best 4 runs from each participating group (0.2187).

Both FILTER-10000 and ENRICH algorithms worked with textual results as primary list, and merge it with all the visual results lists (secondary). TEXT-FILTER uses visual lists as primaries and merges all of them with the textual one (secondary). The bests results corresponded to the experiments which use the Mahalanobis distance and the AND operator.

4 Conclusions and Future Work

In this participation the MAP value obtained for the text-based baseline experiments was 0.2253, higher than the average MAP (0.2187) calculated from the best 4 runs from each participating group.

For the content-based image retrieval, the results have not been very successful. Our results are lower than the best top ten. The most interesting conclusion in that the Mahalanobis distance works better than the Euclidean one, and the best aggregation method is the AND operator. For following editions more low-level features based on local color descriptors and shape descriptors will be included.

Merged results show that the ENRICH algorithm improves very lightly the baseline. This is important taken into account the poor results obtained from the visual retrieval. FILTER-10000 algorithm improves the textual baseline results in terms of precision at low values.

References

1. Arni, T., Clough, P., Sandersin, M., Grubinger, M.: Overview of the ImageCLEFphoto 2008 Photographic Retrieval Task. In: Peters, C., et al. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 500–511. Springer, Heidelberg (2008)
2. Granados, R., Benavent, X., García-Serrano, A., Goñi, J.M.: MIRACLE-FI at ImageCLEFphoto 2008: Experiences in merging Text-based and Content-based Retrievals. In: Working Notes of the 2008 CLEF Workshop, Aarhust, Denmark (September 2008)
3. Grubinger, M., Clough, P., Müller, H., Deselaers, T.: The IAPR-TC12 benchmark: A new evaluation resource for visual information systems. In: International Workshop OntoImage'2006 Language Resources for Content-Based Image Retrieval, held in conjunction with LREC 2006, Genoa, Italy, May 2006, pp. 13–23 (2006)
4. Leon, T., Zuccarello, P., Ayala, G., de Ves, E., Domingo, J.: Applying logistic regression to relevance feedback in image retrieval systems. Pattern Recognition 40, 2621–2632 (2007)
5. Martínez-Fernández, J.L., Villena-Román, J., García-Serrano, A., González-Cristóbal, J.C.: Combining Textual and Visual Features for Cross-Language Medical Image Retrieval. In: Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., Giampiccolo, D. (eds.) CLEF 2005. LNCS, vol. 4022, pp. 712–723. Springer, Heidelberg (2006)
6. Villena-Román, J., Lana-Serrano, S., Martínez-Fernández, J.L., González-Cristóbal, J.-C.: MIRACLE at ImageCLEFphoto 2007: Evaluation of Merging Strategies for Multilingual and Multimedia Information Retrieval. In: Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D. (eds.) CLEF 2007. LNCS, vol. 5152, pp. 500–503. Springer, Heidelberg (2008)