UPMC/LIP6 at ImageCLEFphoto 2009

Sabrina Tollari
Université Pierre et Marie Curie - UPMC Univ Paris 6
Laboratoire d'Informatique de Paris 6 - UMR CNRS 7606 LIP6
104 avenue du président Kennedy, 75016 Paris, France
sabrina.tollari@lip6.fr

October 29, 2009

Abstract

This working note[1] describes the LIP6 runs for the ImageCLEF photo task 2009. Text retrieval is based on Okapi. Visual retrieval is based on HSV histograms. As we think that text modality should be more efficient than content based image retrieval, we use a non symmetric late fusion between text ranks and visual ranks. Finally, we apply two diversity methods based on visual clustering and random permutation. Results show that visual clustering gives better results than text retrieval only when few text information is given.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [**Database Management**]: Languages—*Query Languages*

General Terms

Measurement, Performance, Experimentation

Keywords

Late fusion, Okapi, Visual diversity, Random diversity

1 Features Extraction and Retrieval

1.1 Text Extraction and Retrieval

We apply the same methods to text topics and text data. We first remove standard English stop words[2] as well as common words specific to image retrieval such as: *photo, photography, image, view, show* and *shot*. All capital letters are replaced by minuscules ones. Then we make some typical stemming. We automatically transform all the dates whose format is *21mar95* in *21 march 1995*. We also remove all numbers except the ones corresponding to a year. The word *vaunderzen* is the only word of the topics that newer appears in the text data. So we use the "Did you mean" function of Google with the query *vaunderzen* and the system proposes *vandeurzen*. So we automatically correct the word *vaunderzen* by *vandeurzen* in the topics.

[1] This is a non official working note
[2] http://ir.dcs.gla.ac.uk/resources/test_collections/cacm

Finally, we use Okapi BM25 [2] to match documents and queries. Let w_i be a word and *doc* be a textual document then we calculate the score between a document and a word as:

$$Score(w_i, doc) = \frac{(k+1)tf_i}{k((1-b)+b\frac{dl}{avgdl})+tf_i} \times \log \frac{N - df_i + 0.5}{df_i + 0.5}$$

with tf_i the term frequency of the word w_i in *doc*, df_i the number of documents containing the word w_i, dl the length of the document *doc*, $avgdl$ the average document length in the text collection, N the total number of documents in the collection, k and b two parameters, usually chosen as $k = 1.2$ and $b = 0.75$.

1.2 Visual Extraction and Retrieval

Visual Features We use the same visual features than in [1]. Each image is decomposed into 3 horizontal bands (top, middle and bottom). From each band, a visual vector composed of an HSV histogram (8x3x3) and standard deviations of the H, S and V values, are extracted. Finally, the visual vectors are concatenated and each image is represented by a visual vector of 51 dimensions.

Visual Retrieval We calculate the classical Euclidean distance between the visual vectors of the images in the topics and the visual vectors of the images of the data collection. Then, for each topic, we calculate the arithmetic mean of the visual distances between an image of the collection and all the images of the topic. Finally, the images of the collections are sorted according to their mean distances to the topic images.

1.3 Late Fusion

For the topics of ImageCLEFphoto 2009, we think that text modality should be more efficient than content based image retrieval. So, for each topic, we select the first N images retrieved using text only (see also [4]). Then, for these N images, we average the text rank and the visual rank. Finally, we sort the N images according to their average ranks.

1.4 Promote Diversity

As in [3, 5], we try two methods of diversity. The first method is based on visual clustering, the second method is based on random diversification.

Visual clustering has been studied for a long time. Two approaches are generally proposed: data clustering and space clustering. The first one requires lots of computation time and should be adapted to distribution of the first images ranked by a given query. The second approach, since it is computed independently of the data, is often less effective, but can be applied extremely fast. We choose to cluster the visual space based on the hue dimension of the HSV space. Contrary to [3], this year as images are decomposed into 3 horizontal bands, we only binarize the 8 bin hue histogram of the top band. Each binary vector correspond to a cluster. The number of clusters is 256 (not all are instantiated), a reasonable number for a re-ranking at P20.

We use the visual space clusters to rerank images. For each query, the system browses the first retrieves images from the most relevant to the less relevant and it builds two lists. If an image has the same visual space cluster as an image of the first list, then this image is put in the second list, if not this image is put in the first list. In this way, all images in the first list are in different visual space clusters. When each cluster are represented in the first list then the final list is the concatenation of the first list and the second list. We call this diversification method: DIVVISUTOP.

We also propose to randomly permute the first 40 retrieve images in order to have a point of comparison. We call this naive method of diversification: DIVALEA.

	Query				Modality		Scores			
Name	T	CT	CD	I	T	I	CR10	P10	map	F-measure
FUSION100	X	X	X	X	X	X	**0.5558**	**0.68**	0.35	0.6117
FUSION100BINTOP	X	X	X	X	X	X	0.5497	0.66	0.3	0.5998
FUSION100RAND40	X	X	X	X	X	X	0.5430	0.67	0.35	0.5982
TEXTONLY	X	X	X		X		0.5258	0.65	**0.37**	0.5813
TEXTONLY	X				X		0.4476	0.48	0.31	0.4614
Results for Query Part 1										
TEXTONLY	X	X	X		X		0.5805	**0.85**	**0.42**	0.6892
FUSION100	X	X	X	X	X	X	**0.5982**	0.8	0.41	0.6831
FUSION100RAND40	X	X	X	X	X	X	0.5825	0.77	0.4	0.6640
FUSION100BINTOP	X	X	X	X	X	X	0.5812	0.76	0.33	0.6572
TEXTONLY	X				X		0.4241	0.5	0.31	0.4590
Results for Query Part 2										
FUSION100BINTOP	X	X	X	X	X	X	**0.5181**	**0.56**	0.28	0.5401
FUSION100	X	X	X	X	X	X	0.5134	**0.56**	0.3	0.5375
FUSION100RAND40	X	X	X	X	X	X	0.5034	**0.56**	0.3	0.5302
TEXTONLY	X	X	X		X		0.4710	0.45	**0.31**	0.4613
TEXTONLY	X				X		0.4710	0.45	**0.31**	0.4613

Table 1: Official results. T: Text, I: Image, CT: Cluster Title, CD: Cluster Description, CR10: cluster recall at 10 documents, P10: precision at 10 documents

2 Submitted Runs

Run 1: UPMCLIP6_T_TXT_textonly Text retrieval based on Okapi using only the title of the topic

Run 2: UPMCLIP6_T-CT-CD_TXT_textonly Text retrieval based on Okapi using the title, the cluster title and the cluster description, but without the sentences containing the word "irrelevant"

Run 3: UPMCLIP6_T-CT-CD-I_TXT-IMG_Fusion100 A non symmetric late fusion is realized between the ranks of the first 100 images retrieved using text retrieval (Run 2) and the ranks of the corresponding images retrieved using content based image retrieval on HSV histograms.

Run 4: UPMCLIP6_T-CT-CD-I_TXT-IMG_Fusion100BinTop For each topic, the images of Run 3 are reranked using DIVVISUTOP method.

Run 5: UPMCLIP6_T-CT-CD-I_TXT-IMG_Fusion100Rand40 For each topic, the first 40 images of Run 3 are randomly permuted (DIVALEA method).

3 Experiments

Topics There are 50 topics. The first 25 topics (Query Part 1) are composed of a title and of a variable number of clusters. These topics have a variable number of clusters, ranging from 2 to 10 (mean=4.36). Each cluster is described by a title, a description and an image. The last 25 topics (Query Part 2) are only composed of a title and of 3 images.

Results Table 1 shows the official results. The best results is obtained with the fusion run without using diversity methods. For Query Part 1, the text-based run using cluster title and description gives better results than the the fusion runs. That means that for our system the cluster title and description are more useful than visual information. For Query Part 2, our

method of visual clustering gives better results than text runs, fusion run and fusion run with random permutation. For any case, the last run is always the run using only topic title.

Acknowledgment. This work was partially supported by the French National Agency of Research (ANR-06-MDCA-002 AVEIR project).

References

[1] Ali Fakeri-Tabrizi, Sabrina Tollari, Ludovic Denoyer, and Patrick Gallinari. UPMC/LIP6 at imageclefannotation 2009: Large scale visual concept detection and annotation. In *CLEF working notes 2009*, 2009.

[2] S. E. Robertson, S. Walker, and M. Beaulieu. Experimentation as a way of life: Okapi at TREC. *Information Processing and Management*, 36(1):95–108, 2000.

[3] S. Tollari, M. Detyniecki, A. Fakeri-Tabrizi, C. Marsala, M.-R. Amini, and P. Gallinari. Using visual concepts and fast visual diversity to improve image retrieval. In *Evaluating Systems for Multilingual and Multimodal Information Access – 9th Workshop of the Cross-Language Evaluation Forum, Revised Selected Papers*, LNCS 5706, 2008.

[4] S. Tollari and H. Glotin. Web image retrieval on ImagEVAL: Evidences on visualness and textualness concept dependency in fusion model. In *ACM Conference on Image and Video Retrieval (CIVR)*, pages 65–72, 2007.

[5] S. Tollari, P. Mulhem, M. Ferecatu, H. Glotin, M. Detyniecki, P. Gallinari, H. Sahbi, and Z.-Q. Zhao. A comparative study of diversity methods for hybrid text and image retrieval approaches. In *Evaluating Systems for Multilingual and Multimodal Information Access – 9th Workshop of the Cross-Language Evaluation Forum, Revised Selected Papers*, LNCS 5706, 2008.