Dublin City University at CLEF 2006: Experiments for the ImageCLEF Photo Collection Standard Ad Hoc Task

Kieran McDonald** and Gareth J. F. Jones

Centre for Digital Video Processing & School of Computing
Dublin City University, Dublin 9, Ireland
email: {kmcdon,gjones}@computing.dcu.ie

Abstract. For the CLEF 2006 Cross Language Image Retrieval (ImageCLEF) Photo Collection Standard Ad Hoc task, DCU performed monolingual and cross language retrieval using photo annotations with and without feedback, and also a combined visual and text retrieval approach. Topics are translated into English using the Babelfish online machine translation system. Text runs used the BM25 algorithm, while visual approach used simple low-level features with matching based on the Jeffrey Divergence measure. Our results consistently indicate that the fusion of text and visual features is best for this task, and that performing feedback for text consistently improves on the baseline non-feedback BM25 text runs for all language pairs.

1 Introduction

Dublin City University's participation in the CLEF 2006 ImageCLEF Photo Collection Ad Hoc task adopted standard text retrieval with and without pseudo relevance feedback (PRF), and a combination of text retrieval with low-level visual feature matching. Text retrieval system is based on a standard Okapi model for document ranking and PRF [1]. Experiments are reported for monolingual English and German document retrieval and bilingual searching with a range of topic languages. Topics were translated for cross-language information retrieval using the online Babelfish machine translation engine [2]. Three sets of experiments are reported: first baseline text retrieval without PRF, second text retrieval with PRF, and finally the third set combines text retrieval and visual feature matching.

The results of our experiments demonstrate that PRF improves on the baseline in all cases with respect to both average precision and the number of relevant documents retrieved. Combined text retrieval with visual feature matching gives a further improvement in retrieval effectiveness in all cases.

This remainder of this paper is organised as follows: Section 2 briefly outlines the details of our standard retrieval system and describes our novel PRF method, Section 3 details our submitted runs, Section 4 gives results and analysis of our experiments, and finally Section 5 concludes the paper.

** Now at MSN Redmond, U.S.A.

2 System Description

The introduction of a new collection for the CLEF 2006 Ad Hoc Photo retrieval task meant that there were no existing retrieval test collections to use for system development. We thus used the previous ImageCLEF St Andrew's collection and related experiments on the TRECVID datasets to guide our selection of fusion methods and retrieval parameters for our experiments.

2.1 Text Retrieval

The contents of the structured annotation for each photo (TITLE, DESCRIPTION, NOTES, LOCATION and DATE fields) were collapsed into a flat document representation. Documents and search topics were processed to remove stopwords from the standard SMART list [3], and suffix stripped using the Snowball implementation of Porter stemming for the English language [4] [5]. While the German language topics and documents were stopped and stemmed using the Snowball German stemmer and stopword list [4].

Based on the development experiments, the text feature was matched using the BM25 algorithm with parameters: $k1 = 1.0$, $k2 = 0$, and $b = 0.5$. When using relevance feedback, the top 15 documents were assumed pseudo-relevant and the top scoring 10 expansion terms calculated using the Robertson selection value [1] were added to the original topic. The original query terms were upweighted by a factor of 3.5 compared to the feedback query expansion terms.

2.2 Visual Retrieval

The visual features were matched using the Jeffrey Divergence (a.k.a. Jensen-Shannon distance) matching function [6, 7]. We can interpret Jeffrey Divergence as measuring the efficiency of assuming that a common source generated both distributions – the query and the document.

The following three visual features were used: HSV colour histogram with 16x4x4 quantisation levels on a 5x5 regional image grid, Canny 8 edge + 1 bin for non-edges feature for each region of a 5x5 regional image grid, and a DCT histogram feature based on the first 5 coefficients quantised each into 3 values for a 3x3 regional image grid. More details on these features can be found in [8].

The results for the three visual features were combined using the weighted variant (i.e. linear interpolation) of the CombSUM fusion operator [9]. The scores for each feature were normalised between 0 and 1 and then the weighted sum was calculated for each document across the three features. The weights used for the colour, edge and DCT feature were respectively: 0.50, 0.30 and 0.20.

The results from the two visual examples for each topic were fused using the CombMAX fusion operator, which took the maximum of the normalised scores from each separate visual result list [9]. Scores were first normalised in the separate visual result sets for each topic image to lie between 0 and 1.

2.3 Text and Visual Retrieval Result Combination

Text and visual runs were fused using the weighted CombSUM operator with weights 0.70 and 0.30 for text and image respectively. Scores were normalised to lie between 0 and 1 in the separate text and visual result sets prior to fusion.

3 Description of runs submitted

We submitted two monolingual runs for German and English and cross language runs for both these languages. For English photo annotations we submitted runs for Russian, Portuguese, Dutch, Japanese, Italian, French, Spanish, German and Chinese topics. While for German photo annotations we only submitted runs for French and English topics. For each language pair including the monolingual runs we evaluated three different approaches: text only queries with and without PRF, and a combined visual and text (with text feedback) run for a total of 39 runs submitted.

4 Summary of Experimental Results

Results for our runs are shown in Table 1 From the table we can see that our multilingual fused text and visual submission performed very well, achieving either the second or top rank of submissions for each language pair in terms of Mean Average Precision (MAP). Text feedback increased the text only results in terms of MAP by on average 16.0%. Fusion with visual results increased these results on average by a further 9.2%. Our experiments produced consistent evidence across language pairs that text feedback and fusion with visual features is beneficial in Image Photo search.

Our monolingual English runs with text feedback and fused visual results performed relatively poorly, achieving 8th of 50 submissions in terms of MAP. The increased competition in the more popular monolingual English category relative to the multilingual categories as well as our limited approach in tuning a single parameter selection for all our submitted runs probably accounts for the relative decrease in effectiveness of our approach in the English monolingual category. Our system parameters were tuned for multilingual topics and in future we should tune separately for the monolingual case. For monolingual searches we would expect the effectiveness of the initial query to be higher than in the multilingual case, and therefore both the appropriate feedback and fusion parameters for combining text and visual results may differ significantly compared to the multilingual case. In our case, the initial English monolingual text query under-performed, but was increased by 20.3% by the text feedback approach we employed. This result was further improved by only 4% through fusion with visual results.

Our results consistently show that our fused text and image retrieval submission outperforms our text-only methods. The average relative improvement in MAP was 9.2%, maximum was 18.9% and minimum was 4.0%. The evaluation

Lang-Pair	*FB*	*Media*	*MAP*	*%chg.*	*P@10*	*P@20*	*P@30*	*Rel. Ret.*	*Rank*
Monolingual Runs									
EN-EN	FB	TXT+IMG	**.2234**	4.0%	.3067	.2792	.2628	2205	8/50
	FB	TXT	**.2148**	20.3%	.3050	.2733	.2467	2052	10/50
	·	TXT	**.1786**	·	.2550	.2275	.2100	1806	21/50
DE-DE	FB	TXT+IMG	**.1728**	7.4%	.2283	.2425	.2328	1812	2/8
	FB	TXT	**.1609**	12.7%	.2283	.2300	.2083	1478	3/8
	·	TXT	**.1428**	·	.2283	.2050	.2017	1212	4/8
Multilingual Runs with German Documents									
EN-DE	FB	TXT+IMG	**.1219**	14.8%	.1850	.1750	.1694	1553	1/5
	FB	TXT	**.1062**	18.9%	.1400↓	.1358	.1272	1121	2/5
	·	TXT	**.0893**	·	.1450	.1317	.1211	814	4/5
FR-DE	FB	TXT+IMG	**.1037**	18.9%	.1517	.1467	.1461	1605	1/3
	FB	TXT	**.0872**	28.8%	.1317	.1225	.1167	1192	2/3
	·	TXT	**.0677**	·	.1150	.1083	.0994	898	3/3
Multilingual Runs with English Documents									
CH-EN	FB	TXT+IMG	**.1614**	10.6%	.2267	.2042	.1939	1835	2/10
	FB	TXT	**.1459**	17.0%	.2033	.1908	.1706	1574	3/10
	·	TXT	**.1247**	·	.1783	.1650	.1606	1321	6/10
DE-EN	FB	TXT+IMG	**.1887**	5.8%	.2600	.2575	.2411	2014	1/8
	FB	TXT	**.1783**	10.5%	.2450	.2342	.2117	1807	2/8
	·	TXT	**.1614**	·	.2233	.2025	.1894	1642	3/8
ES-EN	FB	TXT+IMG	**.2009**	6.0%	.2633	.2525	.2467	2118	2/6
	FB	TXT	**.1895**	17.6%	.2583	.2367	.2117	1940	3/6
	·	TXT	**.1612**	·	.2183	.2075	.1978	1719	4/6
FR-EN	FB	TXT+IMG	**.1958**	6.5%	.2483	.2558	.2489	2026	2/7
	FB	TXT	**.1838**	13.3%	.2367	.2308	.2150	1806	3/7
	·	TXT	**.1622**	·	.2167	.2250	.1972	1652	4/7
IT-EN	FB	TXT+IMG	**.1720**	12.8%	.2167	.2200	.2150	2017	2/14
	FB	TXT	**.1525**	15.2%	.2150	.1833	.1711	1780	3/14
	·	TXT	**.1324**	·	.1917	.1650	.1556	1602	9/14
JP-EN	FB	TXT+IMG	**.1615**	10.7%	.2267	.2042	.1939	1848	2/10
	FB	TXT	**.1459**	17.0%	.2033	.1908	.1706	1591	3/10
	·	TXT	**.1247**	·	.1783	.1650	.1606	1321	8/10
NL-EN	FB	TXT+IMG	**.1842**	6.7%	.2467	.2342	.2194	1906	1/3
	FB	TXT	**.1726**	12.5%	.2150	.2042	.1900	1665	2/3
	·	TXT	**.1534**	·	.1817	.1775	.1706	1477	3/3
PT-EN	FB	TXT+IMG	**.1990**	7.8%	.2683	.2625	.2478	2032	2/6
	FB	TXT	**.1846**	13.8%	.2683	.2525	.2228	1824	3/6
	·	TXT	**.1622**	·	.2483	.2217	.1956	1642	5/6
RU-EN	FB	TXT+IMG	**.1816**	7.3%	.2600	.2400	.2156	1982	2/8
	FB	TXT	**.1693**	9.9%	.2500	.2158	.1911	1760	3/8
	·	TXT	**.1541**	·	.2333	.1867	.1694	1609	6/8

Table 1. Retrieval results for our submitted monolingual and multilingual runs for the ImageCLEF Photo 2006 task. Each run is labelled by its Query-Document language pair, whether it used text feedback (FB) and the mediums fused: text only (TXT) or text fused with image search (TXT+IMG). Each run's effectiveness is tabulated with the evaluation measures: Mean Average Precision (MAP), precision at document cut-offs 10 (P@10), 20 (P@20) and 30 (P@30), the number of relevant documents retrieved for each run (Rel. Ret) and the ranking of the results in terms of MAP compared to all other submitted runs to ImageCLEF Photo 2006 task for the respective language pair. The relative increase in MAP (%chg) between text with feedback and without, and between fused visual and text to text alone (i.e. text with feedback) is also listed for each run.

measures are improved for all tested language pairs. This indicates that our fusion approach is stable and produces reliable results. We suspect that our fusion parameters were a bit conservative in the importance given to the visual results for this task. But on the other hand, if we increase the importance of visual results, we may sacrifice some of the stability and consistency of our results. This will be investigated in followup experiments.

Our results also consistently show that our text runs are improved for all language pairs tested when using text feedback compared to without it. This is true for all language pairs and evaluation measures in Table 1, except for precision at a cut-off of 10 documents for English queries against German documents. The decrease in precision is small and insignificant in this case. At an average relative increase in MAP of 16.0%, a maximum of 28.8% and a minimum of 9.9% for the language pairs, the importance of text feedback is well established for text-based search in this task.

The lack of relevant tuning data because of the significant difference between this year and previous years ImageCLEF photo content and descriptions may have led to a less than optimal choice of parameter for fusing visual and text results. Post-ImageCLEF experiments should be able to quantify the improvements that can be made with better tuning or alternative fusion strategies.

5 Conclusions

Our results for the ImageCLEF 2006 photo retrieval task show that fusing text and visual results achieves better effectiveness than text alone. We also demonstrated that PRF is important for improving the effectiveness of the text retrieval model with consistent improvement in results across language pairs. Future experiments will investigate fusion of text and visual features more deeply, since we believe this still has more to offer than we have shown in our current experiments.

References

[1] Robertson, S. E., Walker, S., Jones, S., Hancock-Beaulieu, M. M., and Gatford, M.,: Okapi at TREC-3. In D.K. Harman, editor, Proceedings of the Third Text REtrieval Conference (TREC-3), pages 109-126. NIST, 1995.

[2] Babelfish `http://babelfish.altavista.com/`

[3] *SMART*, `ftp://ftp.cs.cornell.edu/pub/smart/`

[4] *Snowball* toolkit `http://snowball.tartarus.org/`

[5] Porter, M. F.: An algorithm for suffix stripping. *Program* 14:10-137, 1980.

[6] Rao, C. R.: Diversity: Its measurement, decomposition, apportionment and analysis, Sankyha: The Indian Journal of Statistics, 44(A):1-22, 1982

[7] Liu, J.: Divergence measures based on the Shannon entropy, IEEE Transactions on Information Theory, 37(1):145-151, 1991

[8] McDonald, K.: Discrete Language Models for Video Retrieval, Ph.D. Thesis, Dublin City University, 2005.

[9] Fox, E.A. and Shaw, J.A.: Combination of multiple searches, Proceedings of the Third Text REtrieval Conference (TREC-1994), 243-252, 1994.