FIRE in ImageCLEF 2007

Thomas Deselaers, Tobias Gass, Tobias Weyand, and Hermann Ney
Human Language Technology and Pattern Recognition Group
RWTH Aachen University, Aachen, Germany
deselaers@cs.rwth-aachen.de

Abstract

We present the methods we applied in the four different tasks of the ImageCLEF 2007 content-based image retrieval evaluation. We participated in all four tasks using a variety of methods. Global and local image descriptors are applied using nearest neighbour search for the medical and photo retrieval tasks and discriminative models for the object retrieval and the medical automatic annotation task. For the photo and medical retrieval task, we apply a maximum entropy training method to learn an optimal feature weighting from the queries and qrels from last year. This method works particularly well if the queries are very similar as they were in the medical retrieval task.

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 Managment**]: Languages—*Query Languages*

General Terms

content-based image retrieval, image annotation

Keywords

bag-of-visual-words, maximum entropy

1 Introduction

In this work we present our efforts in the four tasks of ImageCLEF 2007. For all of the experiments, the CBIR system FIRE[1] developed in our group was used.

In the following sections we present our efforts in the *Medical Retrieval Task* [10], the *Photographic Retrieval Task* [8], and the *Medical Automatic Image Annotation Task* [10]. Our efforts for the *Object Retrieval Task* are not described here, but in the according overview paper [1].

2 ImageCLEF 2007 Photographic Retrieval Task

The ImageCLEF 2007 Photographic Retrieval Task is described in [8] and the database used is described in [9], here we describe the methods that we applied in the runs we submitted.

[1] http://www-i6.informatik.rwth-aachen.de/~deselaers/fire.html

Table 1: Overview of the submissions to the photographic retrieval task.

run id	w/ text. inf.	trained on	MAP	comment
FIRE	no	no	0.1172	baseline run
RWTH-FIRE-NT-emp	no	-	0.0834	
RWTH-FIRE-NT-emp2	no	-	0.0824	
RWTH-FIRE-ME-NT-20000	no	2006	0.1122	
RWTH-FIRE-ME-NT-1000	no	2006	0.1102	
RWTH-FIRE-emp	yes	-	0.1969	
RWTH-FIRE-emp2	yes	-	0.1913	
RWTH-FIRE-ME-500	yes	2006	0.1974	
RWTH-FIRE-ME-1000	yes	2006	0.1904	
RWTH-FIRE-ME-30000	yes	2006	0.1938	

We submitted a total of nine runs to the photographic retrieval task, five using textual and visual information jointly and four runs using only visual information, furthermore, we provided a visual baseline run to all participants of ImageCLEF shortly after the queries were released.

For these experiments we used the following image descriptors:

- sparse patch histograms [2]
- clustered patch histograms [4]
- local & global colour descriptors from GIFT [14]
- local & global colour descriptors from GIFT [14]
- global texture features [16]
- monomial invariant feature histograms [12]
- relational invariant feature histograms [11]
- Tamura texture histograms [15]
- image thumbnails of 32x32 pixels
- RGB colour histograms with 512 bins

Furthermore, the textual information was available to the retriever in the same manner as described in [6] and also with a pure cosine-matching similarity measure. These features were extracted for all images and then the feature weights were trained according to [7].

As can be seen in Table 1, textual information greatly helps to achieve a much more precise retrieval result, which was to be expected. In the visual-only runs, maximum entropy training also clearly helps to improve the precision. Nevertheless, none of the tuned visual-only runs achieves the precision of our baseline runs, which is probably due to overfitting.

3 ImageCLEF 2007 Medical Retrieval Task

The ImageCLEF 2007 Medical Retrieval Task is described in [10], here we describe the methods we applied.

We submitted a total of ten runs to the medical retrieval task, five using textual and visual information jointly and five using only visual information. Three of the five runs use feature weights that were trained using the maximum entropy method [7] and the other two runs use an empirically determined set of parameters. The trained runs use the topic of 2005, 2006, and 2005 & 2006 jointly respectively to determine the optimal feature weighting.

Table 2 gives an overview of our submissions to the ImageCLEF 2007 medical retrieval task. For all of these experiments the following image descriptors were used [3]:

- image thumbnails of 32×32 pixels
- image thumbnails of 16×16 pixels reduced to 16 colours (which is very similar to the MPEG colour layout descriptor[13])
- colour histograms in RGB space with 512 bins

Table 2: Overview of the submissions to the medical retrieval task.

run id	w/ text. inf.	trained on	MAP
FIRE-NT-emp	no	-	0.0284
FIRE-NT-emp2	no	-	0.0280
FIRE-ME-nt-tr05	no	2005	0.1473
FIRE-ME-nt-tr06	no	2006	0.2227
FIRE-ME-nt-tr0506	no	2005&2006	0.2328
FIRE-emp	yes	-	0.2457
FIRE-emp2	yes	-	0.2537
FIRE-ME-tr05	yes	2005	0.2922
FIRE-ME-tr06	yes	2006	0.3022
FIRE-ME-tr0506	yes	2005&2006	0.3044

Table 3: Results from the combined runs.

	weight for				
run id	FIRE	OHSU	medGIFT	easyIR	MAP
3fire-7ohsu.clef	3	7	0	0	0.0344
3gift-3fire-4ohsu.clef	3	4	3	0	0.0334
5fire-5ohsu.clef	5	3	0	0	0.0327
7fire-3ohsu.clef	7	3	0	0	0.0325
4gift-4fire-2ohsu.clef	4	2	4	0	0.0322
5fire-5easyir.clef	5	0	0	5	0.0256
7fire-3easyir.clef	7	0	0	3	0.0251
3fire-7easyir.clef	3	0	0	7	0.0244
gift-fire-ohsu-easy.clef	1	1	1	1	0.0220
1gift-1fire-8ohsu.clef	1	8	1	0	0.0201

- global texture features [16]
- monomial invariant feature histograms [12]
- relational invariant feature histograms [11]
- Tamura texture histograms [15]

The textual information was included into the experiments as described in [6], we used one textual information retrieval system using only the English texts. These features were extracted for all images and then the feature weights were trained according to [7].

Again, it can be seen that the incorporation of textual increases the retrieval precision dramatically. Maximum entropy training with the 2006 queries is generally better than with the 2005 queries, which is probably due to the greater similarity with the queries of this year. Combining both yields an even higher precision.

3.1 Combined runs with the medGIFT and the OHSU groups

Furthermore, we combined our results with those from the medGIFT group from Geneva and with the OHSU group from Portland, OR. The combinations were done on a submission file basis. That is, the two groups sent us submissions files which they considered to be good runs and then a new score for an image was created by creating a weighted sum of the scores for that particular image from all runs that should be combined. Unfortunately, none of these runs outperforms any of the individual runs which might be due to the combination on the submission file level: if an image is not included in a submission it has a score of 0.0 for that particular run which might have negative influence on the combination.

An overview of the results for the combined runs with the used weighting is given in Table 3.

rank	run tag	score	error rate [%]
6	`RWTHi6-4RUN-MV3`	30.93	13.2
8	`RWTHi6-SH65536-SC025-ME`	32.98	11.9
10	`RWTHi6-SH65536-SC05-ME`	33.21	12.3
11	`RWTHi6-SH4096-SC025-ME`	34.56	12.7
12	`RWTHi6-SH4096-SC05-ME`	34.70	12.4
13	`RWTHi6-SH4096-SC025-AXISWISE`	44.56	17.8

Table 4: Results from the medical automatic annotation run.

4 ImageCLEF 2007 Medical Image Annotation Task

For the medical image annotation task, we applied the same method as last year which is based on the widely adopted assumption that objects in images can be represented as a set of loosely coupled parts. In contrast to former models [4, 5], this method can cope with an arbitrary number of object parts. Here, the object parts are modelled by image patches that are extracted at each position and then efficiently stored in a histogram. In addition to the patch appearance, the positions of the extracted patches are considered and provide a significant increase in the recognition performance.

Using this method, we create sparse histograms of 65536 ($2^{16} = 8^4$) bins, which can either be classified using the nearest neighbour rule and a suitable histogram comparison measure or a discriminative model can be trained for classification. Here, we used a support vector machine with a histogram intersection kernel and a discriminatively trained log-linear maximum entropy model.

A detailed description of the method is given in [2].

We submitted six runs to the medical automatic annotation task [10]. Four of the runs use the method described above using slightly different parameters. The run `RWTHi6-4RUN-MV` is a combination of these runs, where the wild card character for a position (and all succeeding positions on the same axis) is set, if not at least three of the basis-runs agree about the position. The run `RWTHi6-SH4096-SC025-AXISWISE` is the same method as the other runs, but the code is predicted axis-wise.

An overview of our runs together with their ranking in the official results is given in Table 4.

From the results it can be seen that the last run, which tries to use the hierarchy in the first step cannot compete with the methods that use all data for classification at once. However, a slight accuracy improvement is possible if different well-performing runs are combined in a suitable way.

5 Conclusion

From the results of the medical image retrieval task it can be seen that the maximum entropy method for finding feature weights in image retrieval works extremely well if sufficient training data is available and the queries to be processed are similar to those which occur in the training data.

On the other hand, for the photographic retrieval task, the visual baseline run outperforms all tuned settings which is an indicator for overfitting to the training data of the trained runs. This can be due to the training data not being similar enough to this years topics.

The results of the medical annotation task show that using the class hierarchy can lead to a slight accuracy improvement in a second stage but using it in the first stage could not lead to an improved classification performance.

Acknowledgement

This work was partially funded by the DFG (Deutsche Forschungsgemeinschaft) under contract NE-572/6.

References

[1] Thomas Deselaers, Allan Hanbury, and et al. Overview of the ImageCLEF 2007 object retrieval task. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.

[2] Thomas Deselaers, Andre Hegerath, Daniel Keysers, and Hermann Ney. Sparse patch-histograms for object classification in cluttered images. In *DAGM 2006, Pattern Recognition, 26th DAGM Symposium*, volume 4174 of *Lecture Notes in Computer Science*, pages 202–211, Berlin, Germany, September 2006.

[3] Thomas Deselaers, Daniel Keysers, and Hermann Ney. Features for image retrieval – a quantitative comparison. In *DAGM 2004, Pattern Recognition, 26th DAGM Symposium*, number 3175 in Lecture Notes in Computer Science, pages 228–236, Tbingen, Germany, September 2004.

[4] Thomas Deselaers, Daniel Keysers, and Hermann Ney. Discriminative training for object recognition using image patches. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 157–162, San Diego, CA, June 2005.

[5] Thomas Deselaers, Daniel Keysers, and Hermann Ney. Improving a discriminative approach to object recognition using image patches. In *DAGM 2005, Pattern Recognition, 26th DAGM Symposium*, number 3663 in Lecture Notes in Computer Science, pages 326–333, Vienna, Austria, August 2005.

[6] Thomas Deselaers, Tobias Weyand, Daniel Keysers, Wolfgang Macherey, and H. Ney. FIRE in ImageCLEF 2005: Combining content-based image retrieval with textual information retrieval. In *Workshop of the Cross–Language Evaluation Forum (CLEF 2005)*, volume 4022 of *Lecture Notes in Computer Science*, pages 652–661, Vienna, Austria, September 2005.

[7] Thomas Deselaers, Tobias Weyand, and Hermann Ney. Image retrieval and annotation using maximum entropy. In *Evaluation of Multilingual and Multi-modal Information Retrieval – Seventh Workshop of the Cross-Language Evaluation Forum, CLEF 2006*, LNCS, page to appear, Alicante, Spain, September 2007.

[8] Michael Grubinger, Paul Clough, Allan Hanbury, and Henning Müller. Overview of the ImageCLEF 2007 photographic retrieval task. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.

[9] Michael Grubinger, Paul Clough, Henning Mller, and Thomas Deselaers. The iapr benchmark: A new evaluation resource for visual information systems. In *LREC 06 OntoImage 2006: Language Resources for Content-Based Image Retrieval*, page in press, Genoa, Italy, May 2006.

[10] Henning Müller, Thomas Deselaers, Eugene Kim, Jayashree Kalpathy-Cramer, Thomas M. Deserno, Paul Clough, and William Hersh. Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.

[11] Marc Schael. Texture defect detection using invariant textural features. In *DAGM 2001, Pattern Recogntion, 23rd DAGM Symposium*, volume 2191 of *Lecture Notes in Computer Science*, pages 17–24, Munich, Germany, September 2001. Springer Verlag.

[12] Sven Siggelkow, Marc Schael, and Hans Burkhardt. SIMBA — Search IMages By Appearance. In *DAGM 2001, Pattern Recognition, 23rd DAGM Symposium*, volume 2191 of *Lecture Notes in Computer Science*, pages 9–17, Munich, Germany, September 2001. Springer Verlag.

[13] Thomas Sikora. The MPEG-7 visual standard for content description – an overview. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(6):696–702, 2001.

[14] David McG. Squire, Wolfgang Müller, Henning Müller, and Jilali Raki. Content-based query of image databases, inspirations from text retrieval: Inverted files, frequency-based weights and relevance feedback. In *Scandinavian Conference on Image Analysis*, pages 143–149, Kangerlussuaq, Greenland, June 1999.

[15] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki. Textural features corresponding to visual perception. *IEEE Transaction on Systems, Man, and Cybernetics*, 8(6):460–472, June 1978.

[16] Boris Terhorst. Texturanalyse zur globalen Bildinhaltsbeschreibung radiologischer Aufnahmen. Research project, RWTH Aachen, Institut für Medizinische Informatik, Aachen, Germany, June 2003.