XRCE's Participation at Patent Image Classification and Image-based Patent Retrieval Tasks of the Clef-IP 2011

Gabriela Csurka, Jean-Michel Renders and Guillaume Jacquet

Xerox Research Centre Europe, 6 ch. de Maupertuis, 38240 Meylan, France
firstname.lastname@xrce.xerox.com

Abstract.
The aim of this document is to describe the methods we used in the Patent Image Classification and Image-based Patent Retrieval tasks of the Clef-IP 2011 track.

The patent image classification task consisted in categorizing patent images into predefined categories such as abstract drawing, graph, flowchart, table, etc. Our main aim in participating in this sub-task was to test how our image categorizer performs on this type of categorization problem. Therefore, we used SIFT-like local orientation histograms as low level features and on the top of that we built a visual vocabularies specific to patent images using Gaussian mixture model (GMM). This allowed us to represent images with Fisher Vectors and to use linear classifiers to train one-versus-all classifiers. As the results show, we obtain very good classification performance.

Concerning the Image-based Patent Retrieval task, we kept the same image representation as for the Image Classification task and used dot product as similarity measure. Nevertheless, in the case of patents the aim was to rank patents based on patent similarities, which in the case of pure image-based retrieval implies to be able to compare a set of images versus another set of images. Therefore, we investigated different strategies such as averaging Fisher Vector representation of an image set or considering the maximum similarity between pairs of images. Finally, we also built runs where the predicted image classes were considered in the retrieval process.

For the text-based patent retrieval, we decided simply to weight differently the different fields of the patent, giving more weight to some of them, before concatenating the different fields. Monolingually, we then used the standard cosine measure, after applying the tf-idf weighting scheme, to compute the similarity between the query and the documents of the collection. To handle the multi-lingual aspect, we either used late fusion of monolingual similarities (French / English / German) or translated non-English fields into English (and then computed simple monolingual similarities). In addition to these standard textual similarities, we also computed similarities between patents based on the IPC-categories they share and similarities based on the patent citation graph; we used late fusion to merge these new similarities with the former ones.

Finally to combine the image-based and the text-based rankings, we normalized the ranking scores and used again weighted late fusion strategy. As our expectation for the visual expert was low, we used a much stronger weight for the textual expert, than for the visual one. We have shown that while indeed the visual expert performed poorly, combined with text experts the multi-modal system outperformed the corresponding text-only based retrieval system.

Keywords

Multi-lingual and multi-modal Information Retrieval, Patent Retrieval, Image categorization, Fisher Vector

1 Introduction

This year the CLEF IP Track had two image related tasks, the Patent Image Classification and Image-based Patent Retrieval subtasks.

The Patent Image Classification task consisted in categorizing a given patent image into one of 9 pre-defined categories of images (such as graph, table, gene sequence, flowchart, abstract drawing, etc.). Our main aim in participating in this sub-task was to test how our image categorizer performs on this type of images and categories. Therefore, we used our usual low level features, namely SIFT-like local orientation histograms as low level features and on the top of that we built a visual vocabularies specific to patent images using Gaussian mixture model (GMM). This allowed us to represent images with Fisher Vectors [2], and to use linear classifiers to train one-versus-all image modality type classifiers.

The main aim of the Image-based Patent Retrieval subtask was to see how images can help to retrieve patents as relevant or non relevant to a given query patent. Therefore, we first investigated different pure visual systems to test how they can work on such a task. As image representation, therefore we used the Fisher Vector[2] representation as in the case of the Image Classification sub-task. The similarity between images hence was given by the dot product between two Fisher Vectors. However, patents in general have several images. Therefore the problem we had to solve was to design methods that better describe similarities between set of patent images in the context of patent prior art. The different strategies we tested, including methods based on predicted classification scores, are detailed in the Section 3.1.

Even if the main objective of this particular task was to investigate the usefulness of image information in predicting prior art, there are obviously other sources of information that we could exploit jointly in order to find relevant prior art. The most natural one is the textual content and its structure (title, abstract, claims and description). But, as each patent document is associated to some IPC-codes, we could also rely on these codes in order to filter or to refine the ranking of potential prior art documents. Finally, the citation graph could be exploited as well: intuitively, the documents cited in the documents that constitute the prior art of a patent are very likely to be themselves prior art of the patent.

More formally, our general approach was to consider the prior art prediction problem as a ranking problem, where patents in the reference collection are sorted by decreasing order of similarity with the query patent. So, the task amounts to define efficient similarity measures between patents, viewing the patents as multi-modal objects. Indeed, patents have different facets: semi-structured multi-lingual text, images and drawings, citations, IPC-codes, etc. Our main strategy detailed in Section 3.2, simply consisted in defining similarity measures for each facet (or mode) and in using late fusion to combine them.

Finally, to combine visual and text based ranking, we normalized the mono-modal similarity scores and used a weighted late fusion for obtaining the final ranking (see Section 3.3). As our expectation for the visual expert was low, we used a much stronger weight for the textual expert, than for the visual one.

2 The Patent Image Classification Task

As image representation we used the Fisher Vector [2] with the improvement suggested in [3]. These improvements were related to normalizations (with power and L2 norm) and to the image representation with spatial pyramids of the Fisher Vectors. It was shown in [3] that using the power norm (*e.g.* with $alpha = 0.5$) allows to diminish the effect of the background. In the case of document images this is very important, as a big amount of the image portion is simply background (white background). As low level features, we used SIFT-like local orientation histograms. Therefore, we first transformed the binary images into a gray scale image and used a random subset of the training data to build a visual vocabulary (GMM) specific to patent images. This allowed us to represent images with Fisher Vectors which works well with linear classifiers. To train the classifiers, we used our own implementation of the Sparse Logistic Regression (SLR) [1], (i.e. logistic regression with a Laplacian prior).

Table 1. Patent Image Classification: overview of the performances of our different runs.

ID	RUN	EER	AUC	TPR
C1	XEROX-SAS.RUNORH_ROTRAIN	0.041898	0.989322	0.907384
C2	XEROX-SAS.RUNORH	0.059673	0.983019	0.848717
C3	XEROX-SAS.FV_ORH_SP	0.081084	0.917983	0.854289
C4	XEROX-SAS.MEAN_ALL	0.081243	0.906276	0.846716

Results of different runs are shown in Table 1. The baseline method is C2, using the system described above. The main difference between this run and the others are detailed below:

- **C1:** As patent images appeares sometimes rotated, and as our low level features are not orientation invariants, we artificially rotated all training images and added to the training set. Then we used the same system as for C2 but with extended training set. As the results show, we indeed slightly improved the classification performance.
- **C3:** We used a spatial pyramid of Fisher Vectors. However, the classification performance was below the performance of C2. We think that the reason might be on one hand that there is no clear structure (geometry) characterizing the images, on the other the background (white space) proportion in several sub-images became more important.
- **C4:** This run was a late fusion between the other 3 runs. It was rather surprising that it leaded to the worst performance.

Finally, in Table 2 we show the confusion matrix for our best method (C1).

Table 2. The confusion matrix of the run C1. Numbers corresponds to number of images, while in the last column the accuracy is in percentage.

	dra	che	pro	gen	flo	gra	mat	tab	cha	ACC
drawing	317	2	1	0	2	8	0	2	0	95%
chemical structure	1	110	0	0	0	0	0	0	0	99%
program listing	0	0	22	2	0	0	1	1	0	85%
gene sequence	0	0	2	19	0	0	0	3	0	79%
flow chart	3	0	0	0	99	2	0	1	0	94%
graphics	29	0	1	0	0	163	0	1	1	84%
mathematics	0	1	3	0	0	0	121	1	0	96%
table	6	0	1	0	0	3	0	54	0	84%
character (symbol)	0	0	0	0	0	0	0	0	17	100%

3 The Image-based Patent Retrieval Task

The aim of the Image-based Patent Retrieval subtasks was to rank patents as relevant or non relevant given a query patent using both visual and textual information. There were 211 query patents and the collection contained 23444 patents having an application date previous to 2002. The number of images varied a lot, from few images to patents containing several hundred of images. In total we had 4004 images in the query patents and 291,566 images in the collection.

The patents belonged to one of the three IPC sub-classes shown in Table 3. These classes were chosen by the organizers as patent searchers often rely on visual comparison for these patent classes to find relevant prior art. In all our runs, for each topic, the retrieval was done only in the corresponding IPC sub-class.

Table 3. Patent Classes considered in the Image-based Patent Retrieval subtasks.

A43B	CHARACTERISTIC FEATURES OF FOOTWEAR; PARTS OF FOOTWEAR
A61B	DIAGNOSIS; SURGERY; IDENTIFICATION
H01L	SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES

The rest of this section is organized as follows. First we describe our visual retrieval system in section 3.1, then our text retrieval system in section 3.2 , and finally show results of the merged mixed runs in section 3.3.

3.1 Image-based Patent retrieval

As image representation, we used the Fisher Vector representation [2] as for the Image Classification Task. The similarity between images is hence given by the dot product of two Fisher Vectors. However, patents in general have several images, hence the aim here was to analyze which is the best strategy to compare a set of images given in a query patent with another set of images present in a potential prior art patent.

We tested two main strategies. In the first case (MEAN), we considered the average distance between all pairs of images (I1) in the two sets, in the second case (MAX) we considered only the maximum distance of all similarities computed on pair of images (I2).

In addition, we also considered to integrate the image-type classifier described in Section 2. Instead of comparing all pairs of images for both image sets, we restrict the comparison (and, consequently, the computation of the MAX and the MEAN values) to pairs of images that were predicted to belong to the same class. Then, to aggregate the values over the set of possible image classses, in the case of the "MEAN strategy' the average over classes was considered, while in the case of the "MAX strategy", the maximum over classes was kept as similarity between the image sets.

As a third approach, we first discarded all images not predicted as "abstract drawing" and computed the similarity between set of images on the remaining images. The intuition behind this last strategy was that abstract drawings might be the most relevant for patent search, while the similarity between e.g flow-charts or tables might be rather misleading (note that we represent our images with local visual information without any OCR or even strong geometry).

Table 4. Image-based Patent Retrieval: overview of the performances of our different approaches. The performances are all shown in percentages.

Model /strategy	MEAN			MAX		
Classifier	ID	MAP	P@10	ID	MAP	P@10
no used	I1	0.56	0.20	I2	1.84	0.75
all modalities	I3	0.80	0.40	I4	1.84	0.70
only drawings	I5	1.09	0.62	I6	**3.51**	**1.85**

ID	submitted runs	img runs	MAP	P@10
I7	XEROX-SAS.MAXMEANMODAD	I5+I6	1.80	0.82
I8	XEROX-SAS.FVORH_3MAX	I2+I4+I6	**2.50**	**1.00**
I9	XEROX-SAS.FVORH_3MAX3MEAN	I1+I2+I3+I4+I5+I6	1.52	0.60

The results on different approaches (top table) and some of their combinations (bottom table) are shown in 4. We can deduce from this table that the MAX strategy is always better than mean. Furthermore, in the case of patent retrieval, considering only drawings is the best option.

3.2 Text-based Patent retrieval

Patents have different facets: semi-structured multi-lingual text, images and drawings, citations, IPC-codes, etc. Our main strategy simply consisted in defining similarity measures for each facet (or mode) separately and to use late fusion to combine them.

As far as the textual content facet is concerned, there are several issues to be solved, namely how to take the structure of the document into account and how to deal with the multilinguality of the document. We decided simply to weight differently the different fields of the patent, giving more weight to the title (4 times more) and to the abstract (2 times more), before concatenating the different fields. Monolingually, we then used the standard cosine measure, after applying the tf-idf weighting scheme, to compute the similarity between the query and the documents of the collection.

In order to cope with the multilinguality in the collection, we adopted two different strategies. The first one (resulting in similarity measures denoted as $SimText1(q,d)$) consists of a late fusion of the monolingual similarities, giving less weights to non-English parts (weights = 0.1 for German and French monolingual similarities, while the weight = 1 for the textual similarities based on the English parts). The second one (resulting in similarity measures denoted as $SimText2(q,d)$) consists of first "translating" non-English parts into English; the translation is done probabilistically, word by word, using dictionaries automatically extracted from the AC (*Acquis Communautaire*) parallel corpus. If the English version of a field (tile, abstract, claims, description) is absent, but the corresponding field has a non-English content, then the English-translated content of the field is used instead. Eventually, all similarities are computed only for the pivot language, namely English.

The IPC-codes were exploited in the following way. A "taxonomical" similarity measure (denoted as $SimTaxon(q,d)$) is defined between a query patent and a patent of the collection as the proportion of IPC-codes that the query and the target patent share in common (in practice, we used only the codes at the finest level). This taxonomical similarity measure was then combined with the purely textual similarities by late fusion:

$$S(q,d) = SimText(q,d) + \alpha.SimTaxon(q,d) \tag{1}$$

with α =0.7.

We also extracted the graph of citations inside the patent collection, as this information was available in specific fields of the documents. Then we added to the previously computed similarity measure $S(q,d)$ a term that is proportional to the average of $S(q,d')$ over all documents d' that are citing document d, in order to re-inforce the relevance score of "the prior art of the prior art":

$$S'(q,d) = S(q,d) + \beta.\text{avg}_{d' \to d}[S(q,d')] \tag{2}$$

with $\beta = 0.1$ and the symbol $\to$ designing the "cite" relationship.

Practically, our runs rely on the following similarity measures:

- **T1**: $S1(q,d) = SimText1(q,d) + 0.7 \cdot SimTaxon(q,d)$
- **T2**: $S2(q,d) = SimText2(q,d) + 0.7 \cdot SimTaxon(q,d)$
- **T3**: $S3(q,d) = S2(q,d) + 0.1 \cdot \text{avg}_{d' \to d}[S2(q,d')]$

The retrieval performances of these runs are shown in 5. We can see that, while translating the terms in the documents into a same pivot language was helpful, adding the similarities based on citation slightly degraded the NDCG and MAP measures (but increased the P@10).

3.3 Multi-modal Patent retrieval

Finally, to combine visual and text-based rankings, we used a simple weighted score averaging. As our expectation for the visual expert was low, we used a much stronger weight (20) for the textual expert, than for the visual one (1). In the Table 6 we show the performances of our multi-modal runs and in Table 7, we show the late fusion results of the different text and image runs (not only the submitted ones) for better comparison.

Table 5. Performance of our pure text based systems.

ID	submitted runs	MAP	P@10	NDCG
T1	XEROX-SAS.LATEMONO	19.68	6.22	35.83
T2	XEROX-SAS.MT	**20.31**	6.57	**36.45**
T3	XEROX-SAS.MT_CIT	19.28	**6.97**	35.94

Table 6. Performance of our multi-modal retrieval systems.

ID	submitted runs	Image	Text	MAP	P@10	NDCG
M1	XEROX-SAS.3MAX3MEAN	I9	T2	**20.92**	6.72	**36.96**
M2	XEROX-SAS.3MAX3MEAN_LATEMONO	I9	T1	20.07	6.17	36.25
M3	XEROX-SAS.3MAX3MEAN_MT_CIT	I9	T3	19.48	**6.97**	36.18
M4	XEROX-SAS.3MAX_LATEMONO	I8	T1	19.78	6.22	36.96
M5	XEROX-SAS.3MAX_MT	I8	T2	20.61	6.77	36.74
M6	XEROX-SAS.MAXMEANMODAD_MT	I7	T2	20.85	6.67	36.91
M7	XEROX-SAS.MAX_MT_CIT	I2	T3	19.38	**6.97**	36.08

From these tables we can see that the visual information always helped, even using the worst visual run. The gain was relatively small (but significant) about 1% absolut for both MAP and NDCG measures. Surprisingly, the best results multi-modal result was not obtained with the combination of the best image and best text expert, but with a poorer visual expert. Indeed, I5 (see table 4) performed significantly worse than I6, however it was better complementing the text run T1 and T2 than the latter, bringing forward new relevant patents.

Table 7. Performance of our multi-modal retrieval systems.

MAP	T1	T2	T3
I1	19.81	20.95	19.45
I2	20.10	20.87	19.37
I3	20.23	20.87	19.52
I4	19.82	20.64	19.41
I5	20.34	**21.16**	19.42
I6	19.63	20.33	19.37
I7	20.06	20.85	19.42
I8	19.78	20.61	19.38
I9	20.07	20.92	18.48

P@20	T1	T2	T3
I1	6.32	6.67	7.01
I2	6.37	6.82	7.01
I3	6.17	6.72	7.01
I4	6.17	6.77	6.97
I5	6.32	6.67	**7.06**
I6	6.32	6.72	6.92
I7	6.67	6.32	7.01
I8	6.22	6.77	6.97
I9	6.17	6.72	6.97

NDCG	T1	T2	T3
I1	36.06	36.97	36.14
I2	36.27	37.00	36.09
I3	36.36	36.87	36.21
I4	36.05	36.77	36.09
I5	36.45	**37.06**	36.10
I6	35.80	36.47	36.02
I7	36.17	36.91	36.09
I8	36.02	36.74	36.08
I9	36.25	36.96	36.18

4 Conclusion

XRCE participated this year in two sub-task of the Clef-IP, the Patent Image Classification and Image-based Patent Retrieval subtasks. Concerning the patent Image Classification Task we have shown that our visual classification system based on Fisher Vectors built on SIFT like local orientation histograms was able to perfectly work on this task with an EER less than 5% and AUC close to 99%. Concerning the Image-based Patent Retrieval subtask, we proposed different strategies to define similarity between Patents based on images. While visual only systems based on these similarities performs poorly, when combined with text experts they were able to perform better than the corresponding text-only based retrieval system.

Acknowledgments

We would like also to acknowledge Florent Perronnin and Jorge Sánchez for the efficient implementation of the Fisher Vectors computation and Sparse Logistic Regression (SLR) we used in our experiments.

References

1. B. Krishnapuram and A. J. Hartemink. Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *PAMI*, 27(6), 2005.
2. F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *CVPR*, 2007.
3. F. Perronnin, J. Sánchez, and Y. Liu. Large-scale image categorization with explicit data embedding. In *CVPR*, 2010.