UPMC/LIP6 at ImageCLEFphoto 2008: on the exploitation of visual concepts (VCDT)

Sabrina Tollari, Marcin Detyniecki, Ali Fakeri-Tabrizi, Massih-Reza Amini, Patrick Gallinari
Université Pierre et Marie Curie-Paris6
Laboratoire d'Informatique de Paris 6 - UMR CNRS 7606
104 avenue du président Kennedy, 75016 Paris, France
`firstname.lastname@lip6.fr`

Abstract

In this working note, we focus our efforts on the study of how to automatically extract and exploit visual concepts. First, in the Visual Concept Detection Task (VCDT), we look at the mutual exclusion and implication relations between VCDT concepts in order to improve the automatic image annotation by Forest of Fuzzy Decision Trees (FFDTs). In our experiments, the use of the relations do not improve nor worsen the quality of the annotation. Our best VCDT run is the 4th ones under 53 submitted runs (3rd team under 11 teams). Second, in the Photo Retrieval Task (ImageCLEFphoto), we use the FFDTs learn in VCDT task and WordNet to improve image retrieval. We analyse the influence of extracted visual concept models to the diversity and precision. This study shows that there is a clear improvement, in terms of precision or cluster recall at 20, when using the visual concepts explicitly appearing in the query.

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

General Terms

Measurement, Performance, Experimentation

Keywords

Forest of Fuzzy Decision Trees, Cooccurrences Analysis, Visual Concepts, Multi-Class Multi-Label Image Classification, Multimodal Image Retrieval, TF-IDF, Language Model, WordNet

1 Introduction

In this working notes, we comment the runs that were submitted, by the UPMC/LIP6, to the Visual Concept Detection Task (VCDT) and for the Photo Retrieval task (ImageCLEFphoto) of ImageCLEF 2008. For this challenge, we focus our efforts on the study of how to automatically extract and exploit visual concepts.

In the next section, we present our methods and results in VCDT task. In section 3, we describe the techniques we use in the ImageCLEFphoto task, especially how we use the VCDT concepts in this task and our diversification method. Finally, in the last section, we conclude.

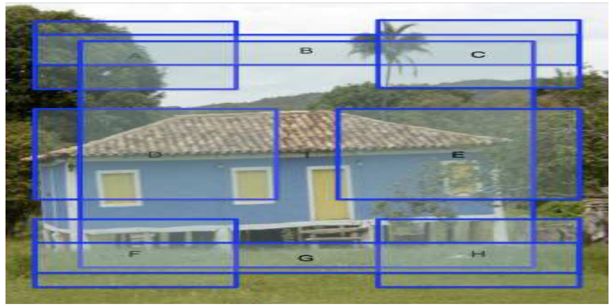


Figure 1: Image segmentation for the extraction of visual descriptors in VCDT task

2 Visual Concept Detection Task (VCDT)

Automatic image annotation is a typical inductive machine learning approach. It has as starting point a set of correctly labeled examples used to train or build a model. In a second stage, the model is used to perform an automatic classification of any forthcoming examples, even if they have not been already met before. Inductive machine learning is a well-known research topic, with a large set of methods, one of the most common being the decision tree approach (DT).

2.1 Forests of Fuzzy Decision Trees (FFDT)

One limitation when considering classical DTs is its robustness and threshold problems when dealing with numerical or imprecisely defined data. The introduction of fuzzy set theory, leading to the fuzzy decision tree approach (FDT), enables us to smooth out these negative effects. In general, inductive learning consists on raising from the *particular* to the *general*. A tree is built, from the root to the leaves, by successive partitioning the training set into subsets. Each partition is done by means of a test on an attribute and leads to the definition of a node of the tree (for more details, see [3]). When addressing unbalanced and large (in terms of dimension and size) data sets, it has been shown in [4] that it is interesting to combine several DTs, obtaining a Forest of Decision Trees. Moreover, when combining the results provided by several DTs the overall score becomes a degree of confidence in the classification. Note that is not the case for the scores provided by a single decision tree.

In the learning step, a forest of FDTs (FFDT) was constructed for each concept X of the VCDT challenge. A FFDT is composed of n FDTs. Each FDT F_i of the forest is constructed based on a training set T_i, each being a balanced random sample of the whole training set.

In the classification step, each image I is classified by means of each FDT F_i. We obtain a degree $d_i(I, X) \in [0, 1]$ for the image to be a representation of the concept X. Thus, for each I, n degrees $d_i(I, X),\ i = 1 \ldots n$ are obtained from the forest. Then all these degrees are aggregated by a simple vote, which mathematically corresponds to the sum of all the degrees: $d(I, X) = \sum_{i=1}^{n} d_i(I)$. Finally, to decide if an image presents a concept or not, we can use a threshold value t (with $0 \leq t \leq n$).

2.2 Cooccurrences analysis

DTs learn each concept independently, but concepts can be related. For instance, when a scene can not be simultaneously *indoor* and *outdoor*, or if we observe that is *overcast*, it implies that we have the concept *sky*. Here, we propose to use cooccurrence analysis to automatically find these relations. Once we have discovered the relation, we need a rule to resolve the conflicting annotations. In fact, each concept is annotated by a FFDT, with a certain confidence degree. For instance, for each image, we will have a degree of having the concept *outdoor* and a certain degree of having *indoor*. We know that both can not appear simultaneously, something has to be done.

We propose to use simple rules. In this paper, we study two type of relations between concepts: exclusion and implication.

Exclusion discovery and rule To discover the *exclusions*, we need to look at what concept *never* appear together. For this, we calculate a cooccurrence matrix COOC. Since there may be some noise (e.g. annotation mistakes), we use a threshold α to decide which pair of concepts never appear together. Once we know which concepts are related, we apply a resolution rule to the scores provided by the FFDT. We choose the rule, that for mutually excluding concepts, eliminates (i.e. gives a confidence of zero) to the label having the lowest confidence. For instance, if we have *outdoor* with a degree of confidence of 42/50 and *indoor* with a degree of 20/50 then we will say that it is certainly not *indoor* and its degree should equal 0. This gives us the following algorithm:

- let COOC be the concept cooccurrence matrix
- for each test image I:
  - let d(I,X) be the FFDT degree of I for concept X
  - for each couple of concepts (A,B) where $COOC(A, B) \leq \alpha$ (*discovery*)

    if d(I,A) > d(I,B) then d(I,A)=0 else d(I,B)=0 (*resolution rule*)

Implication discovery and rule To discover *implications*, we need to look, by definition of the implication, at the cooccurrence of the absence of concepts and of the presence of concepts. The resulting cooccurrence matrix COOCNEG is non symmetric, which reflects the fact that one concept may imply another one, but the reciprocal may not be true. The resolution rule says that if a concept implies another one, the confidence degree of the latter should be at least equal to the former. Since there may be some noise, we use a threshold β to decide which concepts imply other ones. We obtain the following algorithm:

- let COOCNEG be the concept cooccurrence asymmetric matrix between a concept and the negation of an other concept
- for each test image I:
  - let d(I,X) be the FFDT degree of I for concept X
  - for each couple of concepts (A,B) where $COOCNEG(A, B) \leq \beta$ (*discovery*)

    d(I,B)=max(d(I,A),d(I,B)) (*resolution rule*)

2.3 Visual Descriptors

The *visual descriptors* used in this paper are exclusively color based. In order to obtain spatial-related information, the images were segmented into 9 overlapping regions (see figure 1). For each region, we compute a color histogram in the HSV space. The number of bins of the histogram (i.e. numbers of colors) reflects the importance of the region by being valued. The large central region (the image without borders) represents the purpose of the picture. Two other regions, top and bottom, correspond to a spatial focus of these areas. We believe that they are particularly interesting for general concepts (i.e. not objects), as for instance: sky, sunny, vegetation, etc. The remaining regions (left and right top, left and right middle, left and right bottom) are described in terms of color difference between the right and the left. The idea is to explicit any *systematic* symmetries. In fact, objects can appear on either side. Moreover, decision trees are not able to automatically discover this type of relations.

2.4 VCDT Experiments and Results

The VCDT corpus contains 1827 train images and 1000 test images. There are 17 concepts. A train image is labeled in average by 5.4 concepts (standard deviation=2.0, between 0 (2 images) to 11 concepts by image). A concept label in average 584 train images (standard deviation=490,

indoor	outdoor, water, roadorpathway, tree, mountains, beach, buildings, sky, sunny, partlycloudy, overcast
day	night
night	roadorpathway, tree, mountains, beach, sunny, partlycloudy, overcast, animal
roadorpathway	beach
sunny	partlycloudy, overcast, animal
partlycloudy	overcast

Table 1: Automatically discovered exclusive relations. Concepts of the first column are mutually exclusive with *each* of the concepts of the second column

between 68 to 1607 train images by concept). All the forests are composed of 50 trees. This task corresponds to a multi-class multi-label image classification.

Exclusive and implication relations A preliminary step before building our runs is to study cooccurrence values to discover exclusions and implications.

For the 17 concepts, there are 136 cooccurrences values. Those values vary from 0 to 1443 (there are 1827 train images). We set $\alpha = 5$ (two concepts are considered exclusive if at the maximum 5 of the 1827 training images were annotated as presenting the two concepts in the training sets). For the same reason, we set $\beta = 5$ (a concept implies an other concept if at the maximum 5 training images are not annoted by the first concept, but annoted by the second one).

Our system automatically discovered 25 exclusive relations (table 1) and 12 implication relations (table 2). We found not only most of the relations suggested in the schema describing the training data, but also several other ones. For the latter, some are logic and some are the result of the fact that some labels are not very frequent. On table 1, the concepts of the first column are mutually exclusive with *each* of the concepts of the second column (taken individually). We notice, for instance, that *sunny* and *night* never appear together, but also that there is never a *beach* and a *road* together. On table 2, each concept implies the concept in the second column. We found for instance that *tree* implies *vegetation*, but less trivially that *water* implies *outdoor*.

Description of runs In order to understand the effects of the cooccurrence analysis in a concept annotation task, we submitted the following six VCDT runs:

runA B50trees100pc: degrees of confidence as direct results of the FFDT of each concept.

runB B50trees100pc_T25: same as runA, but with a class decision based on a threshold t of 25 (the FFDTs' degrees varying from 0 to 50).

runC B50trees100pc_COOC5: same as runA, filtered by the exclusion resolution rules.

runD B50trees100COOC5T25: same as runC, but with a decision threshold of $t = 25$.

runE B50trees100C5N5: same as runA, filtered by the exclusion and implication resolution rules.

runF B50trees100C5N5T25: same as runE, but with a decision threshold of $t = 25$.

Besides the submitted runs, for the completeness of this study, we calculated the different error rates for:

runw same as runA, filtered by the inclusion resolution rules.

runx same as *runw*, but with a decision threshold of $t = 25$.

runy random degrees.

runz same as *runy*, but with a decision threshold of $t = 25$.

In order to appreciate the effect of the implication and exclusion rules, we look at the results of the submitted runs. Table 3 gives the scores used in VCDT task (i.e. equal error rate: EER

water, roadorpathway, tree, mountains, beach, sunny, partlycloudy, overcast	outdoor
tree	vegetation
sunny, partlycloudy, overcast	sky

Table 2: Automatically discovered implication relations. Each of the concepts in the first column implies the concept in the second

	Excl. rule	Impl. rule	Without class decision			With class decision (t=25)		
				EER(AUC)	EER(AUC) gains %		EER(AUC)	EER(AUC) gains %
FFDT			runA	24.55 (82.74)	-	runB	26.20 (57.09)	-
FFDT	X		runC	27.37 (71.58)	-11 (-13)	runD	28.83 (54.19)	-10 (-5)
FFDT		X	runw	25.66 (82.48)	-5 (0)	runx	27.51 (54.89)	-5 (-4)
FFDT	X	X	runE	27.32 (71.98)	-11 (-13)	runF	28.93 (53.78)	-10 (-6)
Random			runy	50.17(49.68)	-104(-40)	runz	50.26 (24.89)	-48(-56)

Table 3: Results of VCDT task (EER: Equal Error Rate - AUC: Area under ROC curve). *Runw*, *runx*, *runy* and *runz* were not submitted, but were calculated for seek of completeness

and the area under the curve in the ROC space: AUC). Based on these scores, the exclusion and implication rules seem to worsen the results provided by the FFDTs. We believe that this is due to the fact that these scores are not adapted to boolean classification (and our rules provide boolean decisions). The area under the curve and the equal error rate are interesting when the classification is accompanied by a degree of confidence. Moreover, this measure penalize boolean decision over degrees.

Thus, in order to analyze the real effect of the rules (on a decision framework), we propose to use an adapted measure, similar to the EER, the Normalized Score (NS). This score, used in [1], corresponds to: sensitivity+specificity-1. Figure 2 compares the NS varying t. In the case of simple classification (FFDT), the best threshold value is $t = 25$. It corresponds of a full vote of half of the decision trees. If we compare the NS for best threshold, then we observe that rules (exclusion, implication or both) do not improve nor worse the results. We think that this disappointing result may come from the condition of our resolution rules. In fact, if the FFDTs provide high scores for two concepts, it may not be a good idea to choose either one, because actually we do not really no - since there is a strong contradiction. Clearly a further study is needed.

If we tend to annotate the images easily (using a low threshold), then the use of the exclusion may be interesting to clean up and thus improve the results. An explanation is exclusion rules make less error when one of the two degrees $d(I, A)$ and $d(I, B)$ is very low. If we are rather strict in our decision to label an image with a concept (i.e. we have a high threshold t), then using the implication will improve the results. An explanation is implication rules work well when the two degrees $d(I, A)$ and $d(I, B)$ are both high. Overall, the combined use of inclusion and exclusion gives the best results for any threshold. Unfortunately, it does not outperform (just equals) the best results (for $t \leq 25$).

3 Photo retrieval task (ImageCLEFphoto)

3.1 Text retrieval using TD-IDF and Language Model

In ImageCLEFphoto, we use standard TF-IDF model and a language model (LM) as a base line for text analysis. The idea of language model is to estimate the probability of generation of document D for a given query Q, i.e $P(D|Q)$. We suppose that the distribution of documents in corpus is uniform and also that words into a document are independent, then we have: $P(D|Q) = \prod_{q_i} P(q_i|D)$ which declares a unigram model. The fact that each word of a query belongs or does not belong to a document could be used to rewrite the above probability as a multiplication of

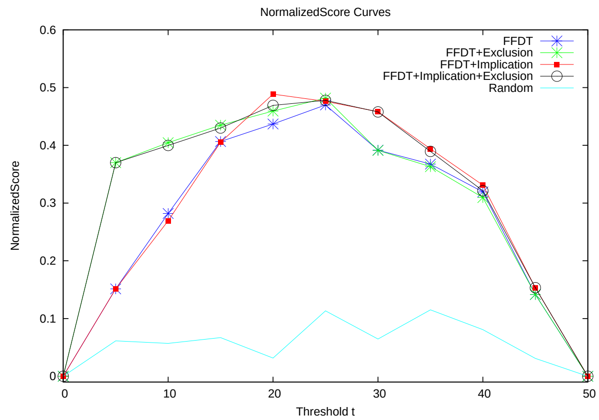


Figure 2: Normalized score for each run (averaged over all concepts)

Query construction using:				LM		TF-IDF	
<title>	<cluster>	<narr>	<narr>-not	P20	Gain %	P20	Gain %
X				0.178	-	0.179	-
X		X		**0.192**	**+8**	0.224	+25
X			X	0.190	+7	**0.260**	**+45**
X	X	X		0.183	+3	0.221	+23
X	X		X	0.190	+7(run1)	0.250	+40(run10)

Table 4: Precision at 20 (P20) in function of the extracted text from topics to build the query. P20 was calculated using the ground truth of ImageCLEFphoto 2007 (not 2008). Only the runs using <title>+<cluster>+<narr>-not were submitted to ImageCLEFphoto 2008 (run1 and run10)

two distribution:

$$P(D|Q) = \prod_{q_i \in D} P_s(q_i|D) \times \prod_{q_i \notin D} P_u(q_i|D).$$

P_s is the model for the words appear in a document and P_u is the model for the words don't appear in a document. $P_u(q_i|D)$ could be estimated by $\alpha_d \times P(q_i|C)$ in which α_d is constant factor and C is the corpus of documents. Thereafter, we can rewrite the formula of $P(Q|D)$ as follows:

$$log\ P(Q|D) = \sum_{q_i \in D} log\ \frac{P_s(q_i|D)}{N(q_i,d)} + |q_i| \times log\ \alpha_d + \sum_i log\ P(q_i|C).$$

$N(q_i, d)$ has a normalization role as IDF in TF-IDF method. Simply, we can consider it as $\beta_d \times P(q_i|C)$ in which β_d is the normalization factor. To calculate $P_s(q_i|D)$, we can use a smoothing method. In our approach, we used Jelinek-Mercer method [7]:

$$P_s(q_i|D) = (1-\lambda)P_{ml}(q_i|D) + \lambda P(q_i|C)\ ,\quad \lambda \in [0,1]$$

in which P_{ml} is the maximum likelihood value.

3.2 Using VCDT concepts in ImageCLEFphoto

Previous works show that combining text and visual information improves image retrieval, but most of this work use an early or late fusion of visual and textual modality. Following the idea of VCDT and ImageCLEFphoto tasks, we propose to use VCDT visual concepts to filter ImageCLEFphoto text runs in order to answer if visual concept filtering can improve text only retrieval.

The difficulty is to determine how to use the visual concepts of VCDT in ImageCLEFphoto 2008. In the VCDT task, we have obtained a FFDT by concept (see sections 2.1 and 2.4). Each

		Filtering by concepts		Diversification					
	Text	VCDT	WN	DIV ALEA	DIV VISU	Modality	P20	CR20	MAP
run1	LM					TXT	0.185	0.247	0.123
run2	LM	X				TXTIMG	0.195	0.257	0.124
run3	LM	X	X			TXTIMG	0.176	0.248	0.122
run4	LM			X		TXT	0.165	0.272	0.112
run5	LM	X		X		TXTIMG	0.154	0.295	0.111
run6	LM	X	X	X		TXTIMG	0.146	0.248	0.114
run7	LM				X	TXTIMG	0.145	0.248	0.090
run8	LM	X			X	TXTIMG	0.151	0.254	0.091
run9	LM	X	X		X	TXTIMG	0.144	0.240	0.090
run10	TF-IDF					TXT	0.250	0.300	0.192
run11	TF-IDF	X				TXTIMG	**0.269**	0.313	**0.194**
run12	TF-IDF	X	X			TXTIMG	0.260	0.293	0.191
run13	TF-IDF			X		TXT	0.227	0.285	0.170
run14	TF-IDF	X		X		TXTIMG	0.232	0.260	0.172
run15	TF-IDF	X	X	X		TXTIMG	0.224	0.254	0.172
run16	TF-IDF				X	TXTIMG	0.204	0.305	0.135
run17	TF-IDF	X			X	TXTIMG	0.214	**0.318**	0.137
run18	TF-IDF	X	X		X	TXTIMG	0.212	0.299	0.133

Table 5: The 18 submitted runs. VCDT filtering: using VCDT visual concepts, VCDTWN filtering: using VCDT visual concepts extended using WordNet, DIVALEA: random permutation, DIVVISU: visual space clustering diversification. All runs are fully automatic (EN-EN-AUTO)

of these FFDTs can give a degree that the corresponding visual concept appears in a new image. In order to make a decision, we put a threshold t to determine if an image contains the given concept according to the corresponding FFDT. First, if the name of a concept appears in the <title> element (VCDT filtering), we propose to filter the rank images list according to the FFDT of this concept. Second, if the name of a concept appears in the <title> element or in the list of synonyms (according to WordNet [2]) of the words in the <title> element (VCDTWN filtering), we also propose to filter the rank images list according to the FFDT of this concept. For example, the <title> of topic 5 is "animal swimming". Using only VCDT filtering, the system automatically determine that it must use the FFDT of the concept *animal*. If, in addition, we use WordNet (VCDTWN filtering), the system automatically determine that it must use the FFDT of the concept *animal* and of the concept *water* (because according to WordNet, the synonym of "swimming" is: "water sport, aquatics").

For each query, we obtain a list of images ranked by their text relevance according to LM or TF-IDF text models. Then, using the decision of the FFDTs, we rerank the first 50 ranked images: the system browses the retrieves images from rank 1 to rank 50. If the degree of an image is lower than the threshold t, then this image is reranked at the end of the current 50 images list.

3.3 Promote Diversity by fastly clustering visual space

For a given query *similar* documents are naturally closely ranked. When a user makes a query, he should want that the first relevant documents are as diverse as possible. So the ImageCLEFphoto 2008 task is very interesting to improve image retrieval, but the definition of diversity in the ImageCLEFphoto 2008 task is not very clear, in particular in term of granularity. In most cases, it is strongly related to the text.

For us, there are two kinds of diversification in the ImageCLEFphoto 2008. The first one is knowledge based: *city, state, country, venue, landmark...*. For this kind of diversification, the use of an ontology (one for country, an other for city...) seems to be a good idea [5], but in real

Text	Visual concept filtering	All 39 topics P20 (gain %)	CR20 (gain %)	Topics modified by filtering Nb topics	P20 (gain %)	CR20 (gain %)
LM	-	0.185 (-)	0.247 (-)	11	0.041 (-)	0.090 (-)
				25	0.148 (-)	0.254 (-)
	VCDT	0.195 (+6)	0.257 (+4)	11	0.077 (+88)	0.126 (+40)
	VCDTWN	0.176 (-5)	0.248 (+1)	25	0.134 (-9)	0.257 (+1)
TF-IDF	-	0.250 (-)	0.300 (-)	11	0.155 (-)	0.161 (-)
				25	0.210 (-)	0.305 (-)
	VCDT	0.269 (+8)	0.313 (+5)	11	0.223 (+44)	0.209 (+30)
	VCDTWN	0.260 (+4)	0.293 (-2)	25	0.226 (+8)	0.294 (-4)

Table 6: Comparison of VCDT and VCDTWN filtering. For VCDT filtering, only 11 topics are modified. For VCDTWN, only 25 topics are modified

application it's hard to determine which ontology applies to a given topic. The second one is based on visual information: *weather condition, group composition, statue....* For this clusters, visual diversification should improve results. Some clusters correspond to both categories: *animal, vehicle type, sport....* For example, for the cluster *animal*, it is possible to distinguish animal in function of the type of animal in the text and also in function of some visual characteristics (like coat, scale, feathers...). As in real applications, it is not obvious to determine automatically which kind of diversification applying for a given query [6], we choose to apply, for all query (even if it is suboptimal), the same kind of diversification (the visual one) by clustering the visual space.

Visual clustering has been studied for a long time now. Two approaches are generally proposed: data clustering and space clustering. The first approach requires lots of calculation time and should be adapted to distribution of the first images ranked by a given query. The second approach, since it is done independently of the data, is often less efficient, but can be applied extremely fast. We choose to cluster the visual space based on the hue dimension of the HSV space. For each image, we binarize its associated 8 bin hue histogram. 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.

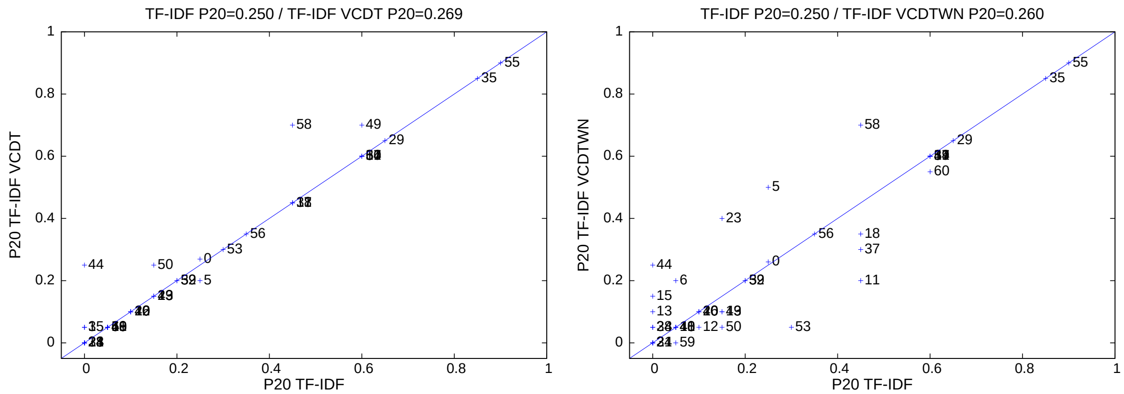
We use the visual space clusters to rerank the 50 retrieve images. For each query, the system browses the retrieves images from rank 1 to rank 50. If an image has the same visual space cluster as an image of highest rank, then this image is reranked at the end of the current 50 images list. In this way, if in the 50 first images, there are n differrent visual space clusters, then at the end of the rerank process, the first n images correspond to strictly different visual space clusters. We call this diversification method: DIVVISU.

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

3.4 ImageCLEFphoto Experiments and Results

The ImageCLEFphoto2008 corpus contains 20k images and 39 topics. Each image is associated with an alphanumeric caption stored in a semi-structured format. These captions include the title of the image, its creation date, the location at which the photograph was taken, the name of the photographer, a semantic description of the contents of the image (as determined by the photographer) and additional notes. In the text retrieval, we use all this elements.

From topics to queries ImageCLEFphoto topics contain different elements: <title>: the title of the topics, <cluster>: defines how the clustering of images should take place, and <narr>: a narrative description of the topics. Table 4 compares different strategies for the construction of text queries in function of the type of text description model (LM and TF-IDF). The results are evaluated using the ImageCLEFphoto 2007 ground-truth. For <title>, <cluster> or <narr>, we use the text of the corresponding element. For <narr>-not, we do not use the sentences of <narr> which contain the word "not". When considering just the <title>, the language model and TF-



(a) TF-IDF versus TD-IDF+VCDT cluster filtering (b) TF-IDF versus TD-IDF+VCDTWN cluster filtering

Figure 3: Details of P20 scores of each topic for TF-IDF and TF-IDF+concept filtering. The point labeled 0 corresponds to the average score of the 39 topics

IDF perform similarly. But, in general, using something more than just the <title> improves the quality (i.e. precision at 20) of results. We observe that TF-IDF has a stronger improvement, first when adding the narrative, and then when filtering the narrative (<narr>-not). According to P20 score, the best combination of elements for query construction in ImageCLEFphoto 2007 (not 2008) is <title>+<narr>-not. However, as we did not look at this scores before we submitted our runs, all the runs we submitted to the 2008 edition are based on the <title>+<cluster>+<narr>-not combination. When we compare the results of *run1* and *run10* in 2007 and in 2008 (table 5), we notice that the P20 results are not the same for *run1* (0.190 in 2007, 0.185 in 2008), but are similar for *run10*. The 2007 ground truth and the 2008 one may be a little bit different.

The submitted runs We submitted 18 runs: 9 based on LM (run 1 to 9, noted Q3 in the name of runs) and 9 based on TF-IDF (run 10 to 18, noted r3tfidf in the name of runs). All use the content of <title>+<cluster>+<narr>-not to construct the query. The results are given in two tables and one figure: table 5 gives the methods used for each submitted run, table 6 compares VCDT and VCDTWN filtering, and figure 4 compares DIVALEA and DIVVISU diversification.

VCDT and VCDTWN filtering To determine if an image should or not contains a visual concept, we choose to set the threshold t to the median of all the degrees values for a given concept (this value varies from 7.3 (*overcast*) to 28.8 (*outdoor*)). We do not use cooccurrence analysis (neither exclusion nor implication rules) in the ImageCLEFphoto task because it was not conclusive in the VCDT task.

Table 6 shows that, for all topics, VCDT filtering improves P20 by 8% and VCDTWN filtering improves P20 by 4% in comparison to TF-IDF P20. Since our method depends on the presence of a concept in the text query, it does not apply to every topic. Using VCDT filtering, only 11 topics where filtered. Using VCDTWN filtering, 25 topics where modified. For the other topics, result images from text retrieval keep the same ranked. Thus, we separate the study into three groups: all the topics, the 11 topics modified by VCDT filtering and the 25 topics for which we applied VCDTWN filtering. On table 6, we observe an improvement on TF-IDF scores of +44% for P20 and +30% for the 11 topics modified by VCDT filtering, but not by VCDTWN filtering (+8% for P20 and -4% for CR20). Figure 3 shows the P20 score for each topic. We notice that, by VCDT filtering, quite all the modified topics are improved, but by VCDTWN filtering, some topics are improved and others are worsened. Then, we conclude that the way we use WordNet is not adapted for this task. Further study is needed.

Diversification Figure 4 compares diversification method scores. DIVALEA and DIVVISU give lower P20 than no diversification, but DIVVISU slightly improves CR20 (in average +2%). So our DIVVISU diversification method works slightly well for diversification, but lowers precision.

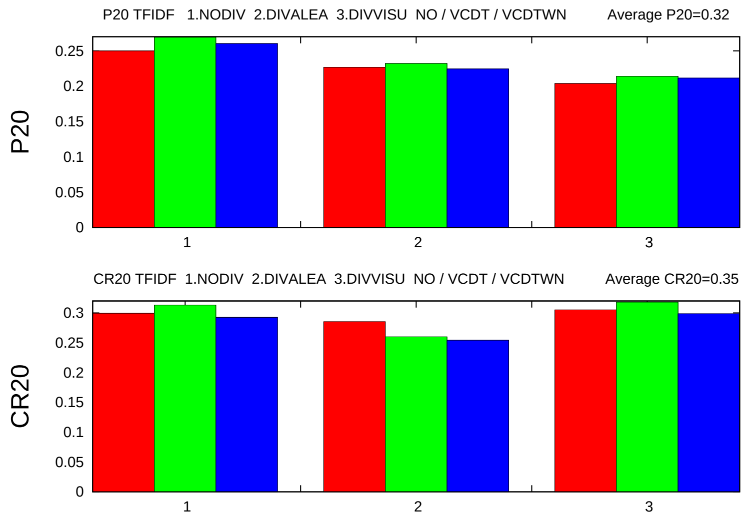


Figure 4: Comparison of diversification methods 1. no diversification, 2. random diversification (DIVALEA) 3. diversification by visual space clustering (DIVVISU). For each diversification method, scores for TF-IDF only (1st bar), TF-IDF+VCDT (2nd bar) and TF-IDF+VCDTWN filtering (3rd bar) are given

4 Conclusion

In this working note, we focus our efforts on the study of how to automatically extract and exploit visual concepts. First, in VCDT task, we look at the mutual exclusion and implication relations between the concepts, in order to improve the automatic labelling. Our best VCDT run is the 4th ones under 53 submitted runs (3rd team under 11 teams). In our experiments, the use of the relations do not improve nor worsen the quality of the labeling. Second, in ImageCLEFphoto task, we analyse the influence of extracted visual concepts models to the diversity and precision, in a text retrieval context. This study shows that there is a clear improvement, in terms of precision or cluster recall at 20, when using the visual concepts explicitly appearing in the query. In our future researches, we will focus on how using image query to improve image retrieval using concept.

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

References

[1] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.

[2] Christiane Fellbaum, editor. *WordNet - An Electronic Lexical Database.* Bradford books, 1998.

[3] C. Marsala and B. Bouchon-Meunier. An adaptable system to construct fuzzy decision trees. In *Proceedings of the NAFIPS'99*, 1999.

[4] C. Marsala and M. Detyniecki. Trecvid 2006: Forests of fuzzy decision trees for high-level feature extraction. In *TREC Video Retrieval Evaluation Online Proceedings*, 2006.

[5] Philippe Mulhem. LIG at ImageCLEFphoto 2008. In *Working Notes of ImageCLEFphoto2008*, 2008.

[6] 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)*, 2007.

[7] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems (TOIS)*, 22(2):179–214, 2004.