TIA-INAOE's Participation at ImageCLEF 2008

H. Jair Escalante, Jesús A. González, Carlos A. Hernández, Aurelio López,
Manuel Montes, Eduardo Morales, Luis E. Sucar, Luis Villaseñor
Research Group on Machine Learning for
Image Processing and Information Retrieval
Department of Computational Sciences
Instituto Nacional de Astrofísica, Óptica y Electrónica,
Luis Enrique Erro No. 1, 72840, Puebla, México
hugojair@ccc.inaoep.mx

Abstract

This paper describes the participation of the INAOE's research group on machine learning for image processing and information retrieval from México. This year we proposed two approaches for the photographic retrieval task. First, we studied the annotation-based expansion of documents for image retrieval. This approach consists of automatically assigning labels to images by using supervised machine learning techniques. Labels are used for expanding the manual annotations of images. Then, we build a text-based retrieval method that uses the expanded annotations. Experimental results give evidence that the expansion could be helpful for improving retrieval performance and diversifying results. However, it is not trivial to determine the best way for combining labels with the other information available. In our second formulation we adopted a late fusion approach to combine the outputs of several heterogeneous retrieval methods. Our aim was to take advantage of the diversity, complementariness and redundancy of documents through ranked lists obtained with different methods and using distinct information. We consider content-based, text-based, annotation-based, visual-concept-based and multi-modal retrieval methods. The fusion of methods achieved competitive performance to that of the best ImageCLEF2008 entries. The heterogeneousness of the retrieval methods proved to be useful for diversifying the retrieval results. For further diversifying the results of our methods we developed a simple strategy based on topic modeling with latent Dirichlet allocation. This technique resulted very helpful for some configurations, though degraded the performance for others. This is mainly due to the quality of the initial retrieval results.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 [**Information Systems and Applications**]: Information Search and Retrieval—*Retrieval models; Selection process; Information Filtering*

General Terms

Performance, Experimentation

Keywords

Multimedia image retrieval, Visual-concept detection, Annotation-based document expansion, Late fusion

1 Introduction

This paper describes the participation of the INAOE's research group on machine learning for image processing and information retrieval (TIA) in the photographic retrieval task of ImageCLEF 2008. This year we submitted a total of 16 runs comprising diverse configurations of the two formulations we adopted. In the first we used automatic image annotation (AIA) methods for expanding the manual annotations of images. Under this formulation a region-level AIA method was used for assigning labels to regions in segmented images. The labels were then combined with the manual annotations of images and the expanded annotations were indexed and queried by using a standard text-based retrieval model. Our assumption is that the labels may provide complementary yet redundant information that can be helpful for improving retrieval performance. One should note that although this method was first proposed by our team for ImageCLEF 2007 [8], in this work we performed experiments with a larger and better training set of annotated regions. Further, we used a different method (in-development) for annotating the images. Experimental results give evidence of slight improvements by using the annotation-based expansion. Interestingly, the diversity of results is increased by using the expanded annotation. These results give evidence that the use of labels generated by AIA methods can be helpful for enhancing retrieval performance. However, it is not trivial to determine the best way for combining labels with the other available information (i. e. image-content and manual annotations).

In our second formulation we considered the late fusion of heterogeneous methods [9]. This approach consists of combining the outputs of independent retrieval methods of diverse nature and based on different sources. Opposed to previous late fusion approaches our formulation considered several retrieval methods per modality, that are different to each other. Our aim was to take advantage of the diversity, complementariness and redundancy of documents through ranked lists of documents obtained with different methods and using distinct information. We considered content-based, text-based, annotation-based, visual-concept-based and multi-modal retrieval methods. A simple weighting scheme allowed us to effectively combine information from diverse sources. Despite the performance of independent retrieval methods is not good, the late fusion approach achieved competitive performance. Further, the heterogeneousness of the retrieval methods proved to be useful for diversifying the retrieval results. We report a few experiments with per-modality and hierarchical fusion, better results were obtained with the latter strategy. Further experiments and a more detailed analysis with this approach are reported elsewhere [9].

The focus of this year photographic retrieval task was on diversifying retrieval results. In order to make varied the results of the late fusion approach we developed a simple strategy based on topic modeling with latent Dirichlet allocation (LDA). The proposed approach consists of finding LDA-topics among the retrieved documents using LDA. LDA-topics can be considered clusters of documents with similar semantic content. Then a single document is selected as representative of each LDA-topic. Representative documents are collocated at the top positions of the ranked list of documents. This technique diversified the results for some configurations of our methods, although it degraded the performance for others. This can be due to the quality of the initial retrieval result. We are currently working in an improved version of the LDA approach for the diversification of retrieval results.

The rest of this paper is organized as follows. In the next Section we briefly introduce the photographic retrieval task. Then, in Section 3, we present the annotation-based approach to image retrieval. Next, in Section 4, we describe the heterogeneous late fusion approach. Next, in Section 5, the LDA approach for diversification of retrieval results is presented. Then, in Section, 6 we report experimental results of our runs. Finally in Section 7 we present conclusions and discuss current and future work directions.

2 Ad-hoc photographic retrieval

This paper presents developments and contributions for the photographic retrieval task of ImageCLEF 2008. The goal of this task is the following: *given an English statement describing an user*

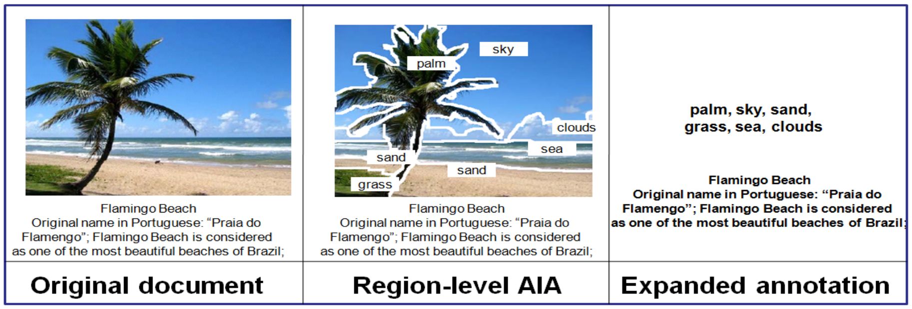

Figure 1: Diagram of the ABDE approach.

information need, find as many relevant images as possible from the given document collection [5, 6]. Organizers provide participants with a collection of annotated images, together with some topics describing information needs. The collection of documents used for *ImageCLEF2008* is the *IAPR TC-12* Benchmark [12]. Each query topic consists of a fragment of text describing a single information need, together with three sample images visually similar to the desired relevant images [6]. Participants use topics content for creating queries that are used with their retrieval systems. Systems runs are then evaluated by the organizers using standard evaluation measures from information retrieval [5, 6]. The focus of this year photographic retrieval task was on diversification of retrieval results. Therefore, organizers encourage participants to submit runs in which the top$-x$ document are both relevant and as diverse as possible. Retrieval methods with explicit mechanisms for diversifying results are supposed to obtain better results. For further information we refer the reader to the respective overview paper [1].

3 Annotation-based document expansion

In this section we describe our AIA-based approach to image retrieval. AIA is the task of assigning semantic labels to images [19]. The main goal of AIA is to allow un-annotated image collections to be searched by keywords. Labels can be assigned either at image-level or at region-level. In the former, labels are assigned to the image as a whole, while in the latter labels are assigned to regions in segmented images. The latter approach can be more useful than the former, because AIA methods can take advantage of spatial context. In this paper we considered region-level AIA methods for expanding manual annotations of images. The annotation-based document expansion (ABDE) approach is depicted in Figure 1.

All of the images in the IAPR-TC12 collection were automatically segmented and visual features were extracted from each region. Using a training set of annotated regions and a multi-class classifier all of the regions in the segmented collection were labeled. For each image, labels were used as the expansion of the original annotation. The expanded annotation was considered as a textual document and a text-based retrieval model was used for indexing the documents. The textual statement in each topic was used as query for the retrieval model. Based on previous work we selected as retrieval engine a vector space model (VSM) with a combination of augmented-normalized term-frequency and entropy for weighting documents [23]. We used the TMG-Matlab^R toolbox for the implementation of all of the text-based retrieval methods [23]. In the rest of this section we provide additional details of the training set we used and the annotation method we considered.

For our experiments with ABDE in ImageCLEF 2007 we faced several issues (due to the training set we used) that made difficult the correct application of ABDE [8]. For this work we considered a better training set composed of about 7000 manually segmented images from the IAPR-TC12 collection. This training set is being created as an effort of INAOE-TIA for providing the community with an extended IAPR-TC12 benchmark that can be useful for studying the use of AIA methods in image retrieval [7]. The images in the training set have been carefully segmented

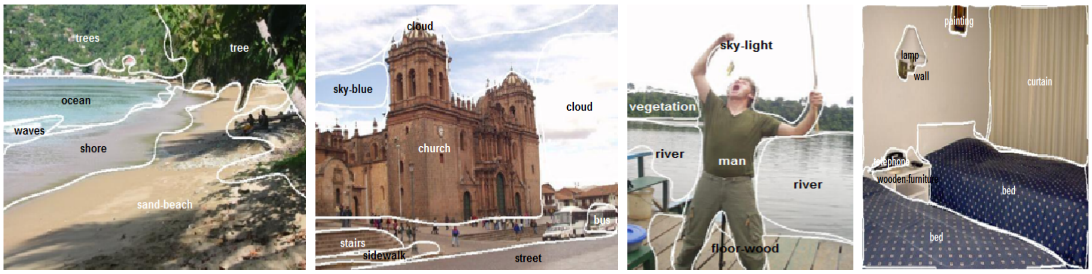

Figure 2: Sample images from the segmented-annotated *IAPR-TC12* collection [7].

following a well defined methodology. Each image is associated with one of 276 labels that have been arranged into a hierarchical structure that facilitates the annotation process. A total of 37,047 regions have been considered for our experiments. Sample images from our training set are shown in Figure 2. The following features were extracted from each region: area, boundary/area, width and height of the region, average and standard deviation in x and y, convexity, average, standard deviation and skewness in both color spaces RGB and CIE-Lab, for a total of 27 features. The training set is therefore composed of features-label pairs.

We used a simple knn classifier as baseline AIA method. Additionally, we considered a recently developed method for improving the quality of annotations [13]. This postprocessing method (referred to as MRFS) is based on a Markov random field that uses spatial relationships between regions for maximizing the coherence of the annotation for each image. The energy function of this random field takes into account a relevance weight obtained from knn and probabilities that reflect the relationships between labels and spatial relationships. For further details we refer the reader to follow the references [13].

4 Late fusion of heterogeneous retrieval methods

In this section we describe our late fusion approach to image retrieval. Late fusion of independent retrieval methods (LFIRM) is one of the simplest and most widely used approaches for combining visual and textual information in the retrieval process [20, 6, 15, 2, 14, 10, 16, 17]. This approach consists of building several retrieval systems (i. e. independent retrieval models, hereafter IRMs) using subsets of the same collection of documents. At querying time, each IRM returns a list of documents relevant to a given query. The output of the different IRMs is then combined for obtaining a single list of ranked documents, see Figure 3. A common problem with this approach is that usually a single IRM is considered for each modality [17, 20, 6, 2, 14, 21, 10, 16]. The latter fact limits the performance of LFIRM because, despite the potential diversity of documents due to the IRMs, there is little, if any, redundance through the IRMs and therefore the combination is not effective [20, 6]. Some LFIRM systems consider multiple IRMs for each modality, however, most of these IRMs are very homogeneous. That is, these methods are variations of a same retrieval model using different parameters or meta-data for indexing [17, 6, 2, 14].

In this work we proposed the combination of heterogeneous IRMs through the LFIRM approach for multimedia image retrieval. We call this approach HLFIRM (heterogeneous LFIRM). Heterogeneousness is important because it can be useful for providing diverse, complementary and redundant lists of documents to the LFIRM approach, reducing the retrieval problem to that of effectively combining lists of ranked documents. For merging the lists we assigned a score to each document in the lists and ranked them in descending order of this score. The combined list was formed by keeping the top-y ranked documents. We assigned a score W to each document d_j in at least one of N lists $L_{\{1,\ldots,N\}}$ of ranked documents as described by Equation (1):

$$W(d_j) = \Big(\sum_{i=1}^{N} \mathbf{1}_{d_j \in L_i}\Big) \times \sum_{i=1}^{N} \Big(\alpha_i \times \frac{1}{\psi(d_j, L_i)}\Big) \qquad (1)$$

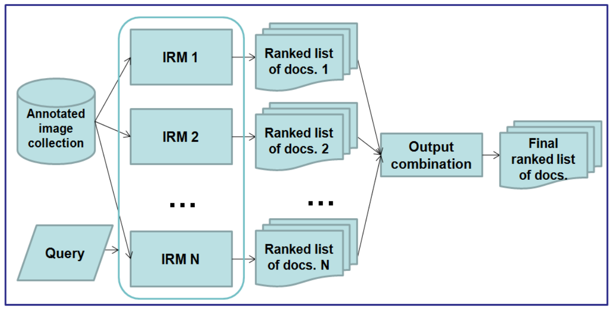

Figure 3: Graphical diagram of the LFIRM approach. The output of different IRMs is combined for obtaining a single list of ranked documents.

ID	Name	Modality	Description
1	FIRE	IMG	CBIR
2	VCDTR-X	IMG	VCDT
3	IMFB-07	TXT+IMG	WQE+IMFB
4	LF-07	TXT+IMG	WQE+LF
5	ABDE-1	TXT+IMG	ABIR
6	ABDE-2	TXT+IMG	ABIR
7	TBIR-1	TXT	VSM t/f
8	TBIR-2	TXT	VSM n/e
9	TBIR-3	TXT	VSM a/g
10	TBIR-4	TXT	VSM a/e
11	TBIR-5	TXT	VSM n/g
12	TBIR-5	TXT	VSM t/g
13	TBIR-6	TXT	VSM n/f
14	TBIR-7	TXT	VSM a/f
15	TBIR-8	TXT	VSM t/e
17	TBIR-9	TXT	VSM t/g

Table 1: List of the IRMs we considered in this work . From rows 7 and on, column 4 describes the local/global weighting schemas for a VSM. Abbreviations are as follows: WQE, web-based query expansion; IMFB, inter-media relevance feedback; LF, Late fusion; t, term-frequency; f, inverse document-frequency; n, augmented normalized term-frequency; e, entropy; a, alternate log; g, global-frequency/term-frequency; l, logarithmic frequency.

where i indexes the N available lists of documents; $\psi(x, H)$ is the position of document x in ranked list H; 1_a is an indicator function that takes the unit value when a is true and α_i ($\sum_{k=1}^{N} \alpha_k = 1$) is the relevance weighting for IRM i. Each list L_i is the output of one of the IRMs we considered, these are shown in Table 1. In the rest of this section we describe these heterogeneous IRMs.

4.1 Image-based IRMs

Two image-based methods were considered for HLFIRM, these are FIRE and VCDTR-X (rows 1 and 2 in Table 1). FIRE is a content-based image retrieval (CBIR) system that works under the query-by-example formulation [11]. FIRE uses the sample images from the topics for querying. Since we are only interested in the output of the IRMs we used the FIRE baseline run provided by ImageCLEF 2007 organizers [6, 11]. In ImageCLEF 2007 the FIRE run we use was ranked at position 377 out of 474.

VCDTR-X is a novel IRM that uses visual-concepts identified in images for retrieval. Visual concepts are indeed labels assigned by image-level AIA methods; the method used for generating the labels is described in [18]. We used the concepts[1] provided by the Xerox Research Center Europe group (XRCE) for building a retrieval model that indexes such concepts. All images (including topic images) were automatically annotated by using this method. The assigned annotations were then used for building a VSM with boolean weighting. Queries for VCDTR-X were the automatic annotations assigned to topic images. No textual information was considered under this formulation. The annotation vocabulary is composed of 17 keywords that describe visual aspects of the images. VCDTR-X is the IRM of the worst performance among those described in

[1] Available from http://www.imageclef.org/2008/iaprconcepts

Table 1, its MAP performance is very close to zero [9].

4.2 Multi-modal IRMs

Four multi-modal IRMs (rows 5-8 in Table 1) of different nature were considered for HLFIRM. ABDE methods are two variants of the method described in Section 3. The first one uses the knn classifier for annotating images, while the second uses the MRFS approach for improving the labeling process. IRMs in rows 3 and 4 of Table 1 are multi-modal methods proposed for the ImageCLEF 2007 competition [6, 8]. IMFB-07 applies inter-media relevance feedback, a technique where the input for a text-based system is obtained from the output of a CBIR system combined with the original textual query [4, 16, 8]. This was our best-ranked entry for ImageCLEF 2007, and for that reason we considered it for this work. LF-07 is an LFIRM run that combines the outputs of a textual method and an CBIR system [8]. The textual-method performs Web-based query expansion, a technique in which each topic-statement is used as a query for *Google*^{R} the top-20 snippets are then attached to the original query [8]. The CBIR system was the FIRE run described in the latter section. This was the run of our group with the highest recall, and that is why we considered for this work. One should note that IMFB-07 and LF-07 were not among the top ranked entries in ImageCLEF2007. These were ranked 41 and 82 in the overall ranked list and achieved a MAP of 0.1986 and 0.1701 respectively. However, in Section 6 we show experimental results that show that these runs resulted very useful for the HLFIRM approach.

4.3 Text-based IRMs

Text-based IRMs (rows 7-17 in Table 1) are variants of a VSM using different weighting schemas. All of these methods index the available text in image annotations by using different weighting strategies (see Table 1). For querying, these methods use the textual statements of topics (including the cluster and narrative fields). We considered ten textual IRMs because, traditionally, textual methods have outperformed both image-based and multi-modal IRMs in past ImageCLEF campaigns [5, 6]. However, the individual performance of textual IRMs is worst than that of the ABDE method [9].

As we can see we have considered a variety of methods that can offer diversity, redundancy and complementariness of documents, opposed to previous work on LFIRM that use single-modality IRMs [17, 20, 6, 2, 14, 21, 10, 16]. These features resulted very useful for HLFIRM that achieved competitive retrieval performance. Further, the use of HLFIRM resulted useful for diversifying retrieval results. All of the IRMs were built by the authors, although some of them were based on methods developed by other research groups [18, 11]. One should note that the individual performance of all of the IRMs we considered is not competitive. Individual IRMs would be ranked at the middle (or near the end) of the overall list of ranked entries for ImageCLEF 2008. However, even with this limitation the best entries with HLFIRM were among the top ranked runs, see Section 6.

5 Diversifying retrieval results

The focus of this year photographic retrieval task was on diversification of retrieval results. Diversity of retrieved documents is important because it facilitates the search process to users. Experimental results (see Section 6) give evidence that HLFIRM is able to diversify retrieval results by itself. However, in order to further increase the variety of documents at the first positions, we developed a diversification approach based on LDA. This approach was applied for each topic as a postprocessing step. We applied this technique to our runs with the late fusion method because this method seemed more promising than the ABDE technique. For each topic we considered the ranked list of documents returned by a retrieval model (in our case we used the output of the HLFIRM approach).

Run	p20	MAP	c20	Avg.	Rel-Ret
Baseline	0.3295	0.2625	0.3493	0.3137	1906
ABDE-Manual	0.3333	**0.2648**	0.351	0.3163	**1913**
ABDE-knn	**0.3397**	0.2554	0.3582	0.3177	1886
ABDE-MRFS	0.3295	0.2546	**0.3733**	**0.3191**	1882

Table 2: Performance of INAOE-TIA entries with ABDE evaluated in ImageCLEF2008. The best result of each measure is shown in **bold**.

LDA is a probabilistic modeling tool widely used in text analysis, image annotation and classification [3, 22]. For text modeling, LDA assumes that documents are mixtures of unknown LDA-topics. LDA-Topics are nothing but probability distributions of words over documents that characterize semantic themes. LDA-Topics are estimated from a collection of documents by using Gibbs sampling or variational inference [3]. Since documents are mixtures of topics we can always calculate the probability of each document given an LDA-topic $P(\mathbf{w}|z_i)$. In this work we associate each document $\mathbf{w}$ to the topic that maximizes the latter probability (i. e. $argmax_i P(\mathbf{w}|z_i)$). In this way, each document is associated to a single LDA-topic, which can be considered a cluster. In this work we used the topic-modeling toolbox due to Steyvers et al. that implements a Gibbs sampling algorithm for inference [22].

For diversifying retrieval results we considered the documents returned by each retrieval model to a query-topic. Considering these documents we used the LDA toolbox for obtaining k LDA-topics (for our experiments we fixed k=20 because the top 20 documents are evaluated in ImageCLEF 2008). Documents were grouped according the LDA-topic they belong to. Then a single document was selected from each LDA-topic as representative of it. The representative document was selected according its relevance weight in the list of ranked documents returned by the retrieval model. The k representative documents were collocated at the top of a new ranked lists of documents. The rest of the documents returned by the retrieval model were collocated below the k documents, in the new list, according their initial relevance weight. In this way diverse yet highly relevant documents are considered at the begining of the ranked lists. Intuitively, a relevant document from each theme (LDA-topic) is put at the top of the new ranked list of documents.

6 Experimental results

In this section we report the results of the runs we submitted to ImageCLEF 2008 for evaluation. A total of 16 runs were submitted comprising diverse configurations of the approaches we adopted. For each configuration we show the precision at 20 documents retrieved (**p20**), mean average precision (**MAP**), cluster recall at 20 documents retrieved (**c20**) and the number of relevant documents retrieved (**Rel-Ret**). We also show the average (**Avg.**) of **p20**, **MAP** and **c20**.

6.1 ABDE entries

First we analyze the performance of ABDE under different settings. The results of our runs with ABDE are shown in Table 2. For all of these entries we used the same weighting schema in the retrieval model, see Section 3. *Baseline* is an VSM that uses only the original image annotations. *ABDE-Manual* is an VSM that expands the original annotations by using the labels from our training set (i. e. only were considered manually assigned labels). *ABDE-knn* uses as expansion the labels of our training set plus the labels assigned with a knn classifier (for those images that have not been manually labeled yet). *ABDE-MRFS* is a run where the labels assigned by knn are further improved with the MRFS approach, see Section 3.

As we can see slightly better results were obtained with the runs that adopted the ABDE approach. The highest MAP and Rel-Ret is obtained by using the manual labels only. While the precision is slightly higher in the *ABDE-knn* run. An interesting result is that the best cluster-recall performance is obtained by the *ABDE-MRFS* entry. The difference is significant with respect to the *Baseline* run. This means that using the labels improved with the MRFS method increases the diversity of results at the first positions, even when the MAP is low. Furthermore, *ABDE-MRFS*

Run	p20	MAP	c20	Avg.	Rel-Ret
All	0.3782	0.3001	0.4058	0.3613	1946
LF-TXT	0.341	0.2706	0.3815	0.3311	1885
LF-VIS	**0.4141**	0.2923	0.3864	0.3642	1966
HLF-EW	0.3795	0.303	0.3906	0.3577	1970
HLF-0.8/0.2	0.391	**0.3066**	0.4033	**0.3667**	**1978**
HLF-0.2/0.8	0.3731	0.2949	**0.4175**	0.3619	1964

Table 3: Performance of INAOE-TIA entries with the late fusion approach that were evaluated in ImageCLEF2008.

is the entry with the highest average performance, offering the best tradeoff between retrieval performance and diversity of results. This result suggest that the labeling improvement due to the MRFS method is indeed useful for improving AIA accuracy. Although, a deeper analysis is required to confirm the latter.

The results shown in Table 2 give evidence that the use of labels generated with AIA methods can be helpful for enhancing the retrieval process. However, the improvements due to the ABDE method are still small. Furthermore, the performance obtained with the best ABDE-entry (i. e. *ABDE-MRFS*) is competitive to methods that only used text for ImageCLEF 2008. However, when compared to multi-modal methods its performance is not that competitive. Therefore, better strategies for combining labels and annotations must be developed. Anyway, despite the mild performance of ABDE methods (when compared to other multi-modal methods) these methods resulted very useful when their outputs were combined with the HLFIRM approach, as described in the next section.

6.2 HLFIRM entries

We performed experiments with several configurations for the HLFIRM approach described in Section 4. First we tried the simple combination of all of the IRMs described in Table 1 (*All*). Then we applied the late fusion approach for combining IRMs that only use text (*LF-TXT*) and IRMs that use images, either image-only or image+text (*LF-VIS*). For these configurations we fixed the relevance weights of IRMs to $\alpha_i = 1$, in this way equal weights are assigned to each IRM, see Equation (1). Then we tried the hierarchical HLFIRM of IRMs. For this configurations we applied the late fusion approach to the already (per-modality) fused *LF-TXT* and *LF-VIS* runs. For the hierarchical fusion we performed experiments with the following weighting schemas. *HLF-EW* is a run that assigns the same weight (i. e. α_i in Equation (1)) to both lists *LF-TXT* and *LF-VIS*. *HLF-0.8/0.2* assigns a weight of 0.8 to the *LF-VIS* list and of 0.2 to the *LF-TXT* run. *HLF-0.2/0.8* assigns a weight of 0.2 to the *LF-VIS* list and of 0.8 to the *LF-TXT* run. The results of all of these configurations are shown in Table 3.

As we can see our results are mixed, though all of the results are very competitive. The highest precision is obtained with the fusion of IRMs that use images either alone or combined with text (*LF-VIS*). This is an interesting result because the individual performance of these methods is poor, see Sections 4.1 and 4.2. Therefore, the HLFIRM approach is effectively combining the outputs of visual IRMs, taking advantage of the diversity and redundancy of results through the individual lists. The performance of text-based methods (*LF-TXT*) is not bad at all. In fact the *LF-TXT* run is among the top-5 ranked entries of methods that only used text. This is also an interesting result because this method is very easy to implement and to apply in practice. It only requires building several text-based retrieval models and combining its outputs by using Equation (1). No complex processes or natural language processing tools were required.

The highest MAP and the best average performance is obtained with by the *HLF-0.8/0.2* entry. This is our best entry in ImageCLEF 2008, ranked among the top-15 runs. This is another interesting result because traditionally in late fusion image retrieval better results are obtained by weighting high text-based methods [17, 20, 5, 6, 2, 14, 21, 10, 16]. This result suggest that, despite their poor performance (see Sections 4.1 and 4.2), the visual-based methods are more helpful for the HLFIRM approach. However, text-based methods are also useful for improving retrieval performance. Cluster-recall is high for most entries in Table 3, giving evidence that the

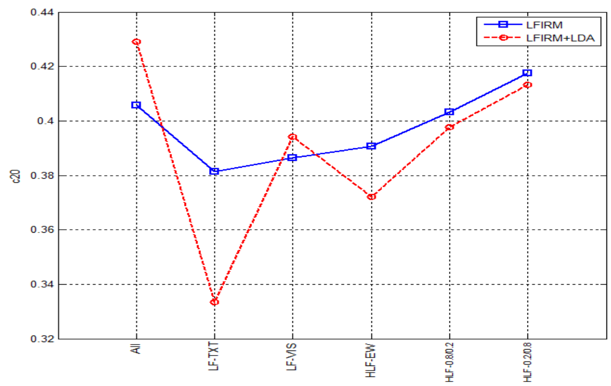

Figure 4: Cluster recall at 20 documents (**c20**) for the HLFIRM entries described in Table 3 with (red dashed-line) and without (blue solid-line) the diversification technique based on LDA.

HLFIRM approach can be useful for diversifying retrieval results. This is due to the fact that with HLFIRM the top ranked documents of each IRM are collocated at the top positions of the fused list, since the IRMs are different in nature, they use to retrieve different documents at the first positions. Therefore, the documents at the first positions in the fused list are diverse. The entry *HLF-0.2/0.8* is ranked 7^{th} in **c20** over entries using text and images, even when we did not have applied any explicit method for diversifying results.

6.2.1 Diversifying results of HLFIRM entries

We applied the LDA approach described in Section 5 as postprocessing to diversify results of the late fusion entries. Results of this experiment are shown in Figure 4, since the impact of the LDA technique is on the diversity of results we only plot the **c20** measure.

As we can see the LDA approach resulted useful only for 2 out of 6 runs (*All* and *LF-VIS*). The best **c20** performance is now achieved with the *All* run. This entry obtained a **c20** of 0.4291, which is ranked at the 4^{th} position in the ranked list of entries that use image and text (in **c20**). However, the average performance of the *All* entry is now of 0.2995. In fact the average performance of all of the runs is decreased by using the LDA approach.

This postprocessing resulted useful for *All* and LF-VIS because for these runs the LDA-topics were effectively identified and the initial ranking was also useful for identifying the representative document of each LDA-topic. For the rest of the entries the LDA approach could not improve the diversity of results. This is due to the quality of the initial list of retrieved documents. Thus even when the LDA approach could effectively find LDA-topics the representative document identified for each LDA-topic is not relevant to the original query. Results obtained with the LDA approach suggest that this technique can be helpful for improving the diversity of retrieval retrieval results. However, it is clear that the average performance of methods decreases by adopting this technique. Therefore, either the initial list of retrieved documents needs to be improved or the LDA technique needs to be modified in order to effectively diversify results for retrieval methods.

7 Conclusions

We have described the participation of INAOE-TIA research group in the photographic retrieval task of ImageCLEF2008. This year we adopted two approaches to the multimedia retrieval problem: annotation-based document expansion (ABDE) and late fusion of independent-heterogeneous retrieval methods (HLFIRM). Further, we proposed a technique based on topic modeling with latent Dirichlet allocation for diversifying retrieval results. Experimental results with the ABDE method give evidence that the use of automatic annotations can be useful for improving retrieval performance. However, the slight improvements in retrieval performance show that either the

ABDE approach may be not the best way for combining labels/annotations or that better annotation methods are required. An interesting result with ABDE is that diversity of results increased significantly by using the labels obtained with our MRFS approach. Showing the annotation effectiveness of MRFS and that the automatic annotations introduced diversity into the retrieval process. These results motivate further research on several directions, including the improvement of image annotation methods, the study of different strategies for combining automatic and manual annotations for multimedia retrieval and analyzing the potential diversity/complementariness/redundancy offered by automatic annotations.

Results with the HLFIRM approach confirm that late fusion is a very useful approach to image retrieval. Despite the individual performance of the methods we considered was not good, our runs with HLFIRM showed competitive performance to that of the best ranked entries in ImageCLEF 2008. An interesting finding is that better results were obtained by assigning a higher weight to visual retrieval methods instead to textual ones. Furthermore, the use of heterogenous methods allowed the HLFIRM approach to implicitly diversify the retrieval results. Current work with the HLFIRM approach consists of using high-performance individual retrieval methods for the fusion and studying different ways to measure the potential diversity/complementariness/redundancy of individual retrieval methods. Our results with the LDA technique show that it may be useful for further diversifying results. However, it is also very possible to damage the performance of the retrieval results with this technique. Nevertheless, our results motivate further research on the diversification technique.

Acknowledgements. We would like to thank the organizers of *ImageCLEF2008* because of their support. This work was partially supported by CONACyT under project grant 61335 and scholarship 205834.

References

[1] T. Arni, M. Sanderson, P. Clough, and M. Grubinger. Overview of the imageclef 2008 photographic retrieval task. In *Working Notes of the CLEF*. CLEF, 2008.

[2] R. Besancon and C. Millet. Merging results from different media: Lic2m experiments at imageclef 2005. In *Working notes of the CLEF 2005*. CLEF, 2005.

[3] D. Blei. *Probabilistic Models of Text and Images*. PhD thesis, U.C. Berkeley, 2004.

[4] Y. Chang and H. Chen. Approaches of using a word-image ontology and an annotated image corpus as intermedia for cross-language image retrieval. In *Working Notes of the CLEF*. CLEF, 2006.

[5] P. Clough, M. Grubinger, T. Deselaers, A. Hanbury, and H. Müller. Overview of the imageclef 2006 photographic retrieval and object annotation tasks. In *Working Notes of the CLEF*. CLEF, 2006.

[6] P. Clough, M. Grubinger, T. Deselaers, A. Hanbury, and H. Müller. Overview of the imageclef 2007 photographic retrieval task. In *CLEF 2007*, volume 5152 of *LNCS*. CLEF, Springer-Verlag, 2008.

[7] H. Jair Escalante, C. Hernandez, J. Gonzalez, A. Lopez, M. Montes, E. Morales, E. Sucar, and L. Villase nnor. The segmented and annotated *IAPR-TC12* benchmark. *Submitted to Computer Vision and Image Understanding*, 2008.

[8] H. Jair Escalante, C. Hernandez, A. Lopez, H. Marin, M. Montes, E. Morales, E. Sucar, and L. Villase nnor. Towards annotation-based query and document expansion for image retrieval. In C. Peters et al., editor, *Proceedings of the CLEF 2007*, volume 5152 of *LNCS*, pages 546–553. CLEF, Springer-Verlag Berlin Heidelberg, 2008.

[9] H. Jair Escalante, C. Hernandez, E. Sucar, and M. Montes. Late fusion of heterogeneous methods for multimedia image retrieval. In *MIR08: Proceedings of the 2008 ACM Multimedia Information Retrieval Conference*, Vancouver, British Columbia, Canada, Forthcomming October 2008.

[10] S. Gao, J. P. Chevallet, T. H. D. Le, T. T. Pham, and J. H. Lim. Ipal at imageclef 2007 mixing features, models and knowledge. In *Working Notes of the CLEF*. CLEF, 2006.

[11] T. Gass, T. Weyand, T. Deselaers, and H. Ney. Fire in imageclef 2007: Support vector machines and logistic regression to fuse image descriptors in for photo retrieval. volume 5152 of *LNCS*. Springer-Verlag, 2008.

[12] M. Grubinger, P. Clough, H. Müller, and T. Deselaers. The iapr tc-12 benchmark: A new evaluation resource for visual information systems. In *Proc. of the Intl. Workshop OntoImage'2006 Language Resources for CBIR*, Genoa, Italy, 2006.

[13] C. Hernández and L. E. Sucar. Markov random fields and spatial information to improve automatic image annotation. In *Proc. of the the 2007 Pacific-Rim Symposium on Image and Video Technology*, volume 4872 of *LNCS*, pages 879–892. Springer, 2007.

[14] R. Izquierdo-Beviá, D. Tomás, M. Saiz-Noeda, and J. Luis Vicedo. University of alicante in imageclef2005. In *Working Notes of the CLEF*. CLEF, 2005.

[15] M. M. Rautiainen and T. Seppdnen. Comparison of visual features and fusion techniques in automatic detection of concepts from news video. In *Proceedings of the IEEE ICME*, pages 932–935, 2005.

[16] N. Maillot, J. P. Chevallet, V. Valea, and J. H. Lim. Ipal inter-media pseudo-relevance feedback approach to imageclef 2006 photo retrieval. In *Working Notes of the CLEF*. CLEF, 2006.

[17] V. Peinado, F. López-Ostenero, and J. Gonzalo. Uned at imageclef 2005: Automatically structured queries with named entities over metadata. In *Working Notes of the CLEF*. CLEF, 2005.

[18] F. Perronin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Proceedings of the 2007 Conference on Computer Vision and Pattern Recognition*, pages 1–8, Minneapolis, Minnesota, US, June 2007. IEEE.

[19] J. Li R. Datta, D. Joshi and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2), 2008.

[20] M. Rautiainen, T. Ojala, and S. Tapio. Analyzing the performance of visual, concept and text features in content-based video retrieval. In *MIR '04: Proc. of the 6th ACM workshop on Multimedia information retrieval*, pages 197–204, New York, NY, USA, 2004.

[21] C. Snoek, M. Worring, and A. Smeulders. Early versus late fusion in semantic video analysis. In *Proc. of the 13th Annual ACM Conference on Multimedia*, pages 399–402, Singapore, 2005. ACM.

[22] M. Steyvers and T. Griffiths. *Latent Semantic Analysis: A Road to Meaning*, chapter Probabilistic topic models. Laurence Erlbaum, 2007.

[23] D. Zeimpekis and E. Gallopoulos. Tmg: A matlab toolbox for generating term-document matrices from text collections. In *Recent Advances in Clustering*, pages 187–210. Springer, 2005.