

TIA-INAOE's Participation at ImageCLEF 2007

H. Jair Escalante, Carlos A. Hernández, Aurelio López, Heidy M. Marín,
Manuel Montes, Eduardo Morales, Luis E. Sucar, Luis Villaseñor

Coordinación de Ciencias Computacionales
Instituto Nacional de Astrofísica, Óptica y Electrónica,
Luis Enrique Erro No. 1, 72840, Puebla, México

{hugojair,carloshg,allopez,hmarinc,montesg,emorales,esucar,villasen}@ccc.inaoep.mx

Abstract

This paper describes the participation of *INAOE's* research group on machine learning, image processing and information retrieval (*TIA*) in the photographic retrieval task at *ImageCLEF2007*. Many experiments were performed comprising most of the query and target languages proposed for this years competition. A web-based query expansion technique was proposed for introducing context terms in the query topics. An *intermedia* blind relevance feedback technique was adopted in some runs for query refinement. Furthermore, experiments were performed with a novel technique for query/document expansion based on automatic image annotation methods. Initial experimental results give evidence that this idea could help to improve effectiveness of retrieval methods, though several issues should be addressed.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [Database Management]: Languages—*Query Languages*

General Terms

Measurement, Performance, Experimentation

Keywords

Image retrieval, automatic image annotation, web-based query expansion, intermedia feedback

1 Introduction

Technology innovations have provided humans with more information than we are able to manually analyze. Currently, many huge repositories with textual, visual and auditive information are available. In addition there are collections of mixed data, for example images accompanied with textual descriptions. These last type of collections are referred to as annotated collections (one data type, *text*, is the annotation of another data type, *images*). Since documents in annotated collections are described in two modalities, document retrieval from these collections is a more challenging task than that of traditional information retrieval. This due to the problem of the lack of correspondence between modalities (the *semantic gap*) [11]. Though the consideration of two modalities could help improving accuracy of information retrieval systems. This paper is concerned with the task of retrieval of images in (manually) weakly annotated collections. Weakly annotated image collections are image databases in which each image is accompanied by a very small textual description, regarding the visual or semantic content of the image.

Current solutions for image retrieval from annotated collections are based on a single modality (text or image). Though retrieval accuracy of such systems is limited, mainly, because of the omitted modality and

the querying complexity. In consequence there exist an increasing interest on the development of methods that can take advantage of all of the available information, that is *text + image*. In this research line, an effort is being carried out by the organizers of ImageCLEF¹; the cross-language image retrieval track of the CLEF² forum, whose goal is to investigate the effectiveness of combining text and image for retrieval, to collect and provide resources for benchmarking image retrieval systems and to promote the exchange of ideas which will lead to improvements in the performance of retrieval systems. Organizers provide participants with image collections, query topics and relevance assessments of retrieval systems' runs [9]. This sort of forums are very useful because they motivate research advances on several problems related to information retrieval, furthermore they contribute to the creation of new multidisciplinary research groups and collaborations between participants.

This paper describes the participation of the INAOE's research group on machine learning, image processing and information retrieval (*TIA*) in the photographic retrieval competition at *ImageCLEF2007*. This is the first time *TIA* participates on *ImageCLEF*. A total of 95 runs were submitted comprising all of the target languages and the following query ones: English, Spanish, German, Italian, French, Swedish, Japanese, Portuguese, Russian and Visual. Experiments were carried out with different techniques, like intermedia feedback and fusion of independent retrieval models. Other experiments were carried out with an approach based on web query expansion and with a new method based on automatic image annotation. We propose a web-based technique for expanding *ImageCLEF2007* queries, with which promising results were obtained. Furthermore, we propose a new research direction not explored yet for image retrieval from annotated collections. The approach consist of using automatic image annotation methods for obtaining text from the visual content of the images. This text is then combined with the original annotations of the images (and/or with the queries) and then standard strategies are adopted for retrieving documents. Experimental results show that runs based on text and image improve those based on text only. The best results were obtained with a combination of intermedia feedback and our Web-based query expansion technique. Relatively good results were obtained with the annotation based expansion. However, several issues should be addressed in order to obtain better results with this technique.

The rest of the document is organized as follows. In the next section we briefly introduce the ad-hoc photographic retrieval task, for further information we refer the reader to the overview paper by Grubinger et al [9]. Next, in Section 3 the text-only methods are described. In Section 4 two techniques that were used for combining textual and image information are presented. Then, in Section 5 the proposed annotation-based expansion technique is introduced. In Section 6 the results of our runs are presented and analyzed. Finally, in Section 7, we highlight some conclusions and discuss future work directions.

2 Ad-hoc photographic retrieval

The *ImageCLEF2007* is a track running as part of the *CLEF* campaign. It comprises four tasks related to image retrieval, namely: ad-hoc photographic retrieval, object retrieval, medical image retrieval and automatic annotation for medical images. This paper presents developments and contributions for the first task, that of retrieving images from a collection of annotated photographs. The goal of this task is the following: *given a multilingual statement describing an user information need, find as many relevant images as possible from the given document collection* [4, 9]. Organizers provide participants with a collection of annotated images, together with some topics describing information needs. Participants use these resources with their retrieval systems and submit to the organizers the identifiers of the relevant documents (according to the retrieval system used) for each topic. Organizers evaluate the results set of each submission for every participant and rank submissions according to a standard evaluation measure. The collection of documents used for *ImageCLEF2007* is the recently created *IAPR TC-12* Benchmark [10]. Each query topic consist of a fragment of text describing a single information need, together with three sample images visually similar to the desired relevant images [9]. Participants use topic's content for creating queries that are used with their retrieval systems. System's runs are then evaluated by the organizers using standard evaluation measures like *MAP* [9].

¹<http://ir.shef.ac.uk/imageclef/2007/>

²<http://clef-campaign.org/>

3 Textual methods for image retrieval

The predominant approach for image retrieval from annotated collections consist of using text-only retrieval methods. This mainly because in some collections annotations describe effectively the visual and semantic content of images, for example the collection used for *ImageCLEF2006*. However in realistic collections images are described by only a few words indicating the semantic content of each image, for example the collection used for *ImageCLEF2007*. In these scenario information extracted from the images could be helpful for improving retrieval performance. For this reason we performed experiments with methods based on both text and images, text-only runs were also submitted for measuring the gain we have by using visual information. Independently of the type of run all of our submissions used a textual retrieval system (for the mixed runs text was extracted from images and used with the textual retrieval system), which is described in Section 3.1. Furthermore, in Section 3.2 the proposed technique for expansion of textual queries is presented.

3.1 Baseline retrieval system

Since text was used in all of the *TIA*'s runs a textual retrieval system was implemented. For this purposed the *TMG* Matlab^R toolbox was used, kindly provided by Zeimpekis et al [17]. This toolbox includes standard methods for indexing and retrieval of middle size text collections. We decided to use this method because working with Matlab^R allows the easy processing of images (through predefined functions), necessary for the development of the annotation based expansion technique.

After removing meta-data and useless information, the text of the captions was indexed on the four target languages (English, Spanish, German and Random), resulting on four indexed collections. For indexing we used a *tf-idf* weighting, English stop words were removed and standard stemming was applied [1, 17]. Queries for the baseline runs were created by using the text in topics as provided by the organizers, after removing meta-data. For multilingual experiments queries were translated using the online Systran³ translation software. For retrieval we considered the cosine similarity function [1].

3.2 Web based query expansion

The Web is the largest repository of information ever existed. There are millions of documents available in the Web that can be used to extract semantic knowledge. In this work we propose a web-based query expansion technique in order to incorporate context terms into the original queries, with the hope that expanded queries are able to reach (through a retrieval model) relevant documents containing other terms than the ones contained in the original queries.

The proposed approach is a very intuitive method that uses the Google^R search engine. For each topic, we take the textual description and submitted a search to Google^R; the top-*k* snippets returned by the search engine are considered for expanding a query. We tried two approaches that we called *naive* and *repetition*. The naive approach consist of taking the snippets as they are returned by Google^R with no preprocessing. On the other hand, the repetition approach consist of retaining the terms that most co-occurred among the snippets. For the experiments reported here the top-20 snippets were considered. Queries expanded with the naive method resulted on larger queries, while the ones expanded with the repetition method included only a few extra terms.

4 Mixed methods for image retrieval

In order to consider both textual and visual information in the retrieval process, two strategies were adopted: intermedia feedback and late fusion of independent retrieval systems. We decided to use these techniques due to their simplicity, and because both methods have obtained relatively good results on the retrieval task [2, 4].

³<http://www.systranbox.com/>

4.1 Intermedia pseudo-relevance feedback

Relevance feedback is a way of allowing user interaction in the retrieval process. It was first proposed for retrieval from text collections and then used on the content-based image retrieval (*CBIR*) task [15]. Although user interaction is allowed on the retrieval process, always is preferable to develop fully automatic systems. Pseudo-relevance feedback (*PRF*) is a variant that does not need user interaction, as it considers as relevant the set of the top- k documents returned by the retrieval system. Queries are refined considering such a set of documents as relevant for the user’s information needs. This technique has been widely used in several text-based image retrieval systems within the *ImageCLEF* [5, 4, 9].

A novel technique based on *PRF* has been proposed for image retrieval from annotated collections [3, 2, 14, 12]. This technique called intermedia feedback (*IMFB*) consist of using a *CBIR*⁴ system with a query image for retrieving documents. The top- k documents returned are assumed to be relevant and the captions of such documents are combined to create a textual query. The textual query is then used with a pure textual retrieval system, and the documents returned by such a system are returned to the user. This technique was first used by the *NTU* group at *ImageCLEF2005* and since then several groups have been adopted it [2, 12, 14]. We decided to perform experiments with this technique in order to improve retrieval effectiveness of the query expansion method proposed, by taking into account information extracted from images. Combined runs of query expansion and *IMFB* consist of applying the query expansion technique to the textual topics. Next, the expanded queries are combined with the captions of the top- k relevant documents according to a *CBIR* system, and then used with our text-only retrieval system. *FIRE* was used as *CBIR* system; using the baseline run provided by the organizers [9].

4.2 Late fusion of independent systems

For some runs we adopted a late fusion strategy for merging the output of independent visual and textual systems. This merging method consists of running two independent retrieval systems using a single (different) modality each. Then the relevant documents returned by both systems are combined.

For this work we proposed the following fusion strategy. Assuming that T_R is the list of relevant documents (ranked in descending order of relevance) according to a textual retrieval system applied to the images’ captions; similarly, V_R is the list of ranked relevant documents according the a *CBIR* system that only uses the images. We combined and re-ranked the documents returned by both retrieval systems, generating a new list of relevant documents $LF_R = \{T_R \cup V_R\}$; where each document $d_i \in LF_R$ is ranked according to the score formula given by Equation (1)

$$score(d_i) = \frac{\alpha \times R_{T_R}(d_i) + (1 - \alpha) \times R_{V_R}(d_i)}{1_{T_R}(d_i) + 1_{V_R}(d_i)} \quad (1)$$

where $R_{T_R}(d_i)$ and $R_{V_R}(d_i)$ is the position in the ranked list of document d_i according to the textual and visual retrieval systems, respectively. $1_{T_R}(d_i)$ and $1_{V_R}(d_i)$ are indicator functions that take the value 1 if document d_i is in the list of relevant documents according to the textual and visual retrieval systems respectively, and zero otherwise. The denominator accounts for documents appearing in both lists of relevant documents (T_R and V_R). Documents are sorted in ascending order of their score. Intuitively with this score documents appearing in both sets (visual and textual) will appear at the top of the ranking, regarding their position on the independent lists of relevant documents. We tried several values for α and the best results were obtained with $\alpha = 0.9$.

5 Annotation-based document and query expansion

The task of automatic image annotation (*AIA*) consists of assigning textual descriptors (words, labels) to images, starting from visual attributes extracted from them. There are two ways of facing this problem, at image level and at region level. In the first case, labels are assigned to the entire image as an unit, not specifying which words are related to which objects within the image. In the second approach, which can

⁴Note that the process can start from text, obtaining query images for a *CBIR* system, as well.

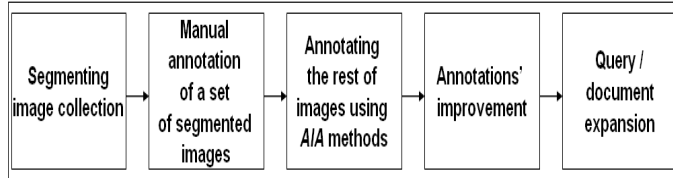


Figure 1: Graphical schema of the proposed approach for *AIA*-based expansion.



Figure 2: Sample images from the *IAPR TC-12 2007* collection, segmented with the normalized cuts algorithm. Manual annotations are shown for each region.

be conceived as an object recognition task, the assignment of labels is at region level; providing a one-to-one correspondence between words and regions⁵. The last approach can provide more semantic information for the retrieval task, although it is more challenging than the former.

In this work we decided to use region-level *AIA* methods for obtaining text from images and then using these textual labels for expanding topics and images' annotations. We believe that by taking both modalities to a common representation and then using standard retrieval strategies can be helpful for improving single-modality approaches. A graphical description of the proposed approach is shown in Figure 1. As we can see the process involves several steps that can be summarized in the following tasks: segmentation, annotation, and expansion.

5.1 Segmentation

The first step for obtaining words from regions within images is obtaining regions from images. This task is known as segmentation and consists of discovering partitions within a given image with the restriction that each of the partitions contains a single object. Many segmentation methods have been proposed, although this is still an open problem in vision. For segmenting the *IAPR-TC 12 Benchmark* collection we decided to use a state of the art algorithm named normalized cuts [16]. This is the algorithm used by most of the region-level annotation approaches [6, 11]. In Figure 2 sample images segmented with normalized cuts are shown. As we can see the algorithm works well for some images, isolating single objects; however for other similar images segmentation is not as good, partitioning single objects into several regions.

Using a recently developed segmentation-annotation interface [13], the full *IAPR-TC 12 Benchmark* collection was segmented with the normalized cuts algorithm. Then visual attributes were extracted from each region. Attributes include color, texture and shape information of the region, for a total of 30 attributes. Each region is then described by its vector of attributes. In order to facilitate reading hereafter we refer to the attribute's vector describing a region simply by the term region.

⁵Note that with region-level annotation we have also an annotation at image level, which is the set of labels assigned to the regions within the image.

1	2	3	4	5	6	7	8	9	10	11	12
Sky	Person	Building	Trees	Clouds	Grass	Water	Mountain	Sand	Other	Furniture	Road
13	14	15	16	17	18	19	20	21	22	23	24
Animal	Snow	Rock	Sun	Vehicle	Boat	Church	Tower	Plate	Flag	Statue	Prize

Table 1: Vocabulary of labels considered for the annotation process. The number above each label is the identifier for such label used in the histogram of Figure 3.

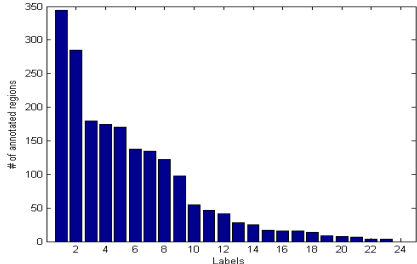


Figure 3: Histogram of regions in the training set annotated with the words defined in the vocabulary.

5.2 Annotation

AIA methods start from visual attributes extracted from the image or regions and assign labels based on some training examples. Each training example is a pair composed of a region and its corresponding label. In order to create a training set of region-label pairs an annotation interface was used [13]. Such interface allows the manual annotation, at region-level, of segmented images, as well as options for re-segmentation and for joining adjacent regions containing the same object. The *IAPR-TC 12 Benchmark* consist of 20,000 images, taking the larger 5 regions for each image we have 100,000 regions to annotate. In consequence we would need a large training set of annotated images-regions. However time constrains allowed us to create only a small training set. For creating the training set we randomly selected around 2% of the total number of regions and manually annotated them using the developed interface.

The set of labels that can be assigned to regions (that is, the annotation vocabulary) was defined subjectively by the authors, by looking at the *ImageCLEF2007* textual topic descriptions. The vocabulary of allowed words is shown in Table 1, and the number of regions, in our training set, annotated with each label is shown in Figure 3. Some labels are supposed to represent several concepts, for example the label *water* was used to label regions of rivers, ocean, sea, and streams. While other labels represent specific objects, such as *swimming-pool* and *tower*. From Figure 3 we can see that there are several labels that have many training examples (for example, *Sky*, *Person*), though several other labels have only a few. This fact together with poor segmentation made difficult the process of annotation. Sample images with their corresponding annotations are shown in Figure 2.

The training set of region-label pairs is used with a *knn* classifier for annotating the rest of un-annotated images-regions on the *IAPR-TC 12 Benchmark* collection for document expansion, or the topic’s images for query expansion. Note that the training set size is very small⁶ for achieving good results with the *knn* algorithm; even when *knn* have obtained good results on this task, outperforming other state of the art annotation methods [7, 8]. In order to overcome in part the issues of poor segmentation and imbalanced/small training set we decided to apply a postprocessing to *knn* for improving annotation accuracy. Recently a method, called *MRFI*, for improving accuracy on *AIA* has been proposed [8]. *MRFI* considers a set of candidate labels for each region and selects a unique label for each region based on semantic information between labels and the confidence of the *AIA* method on each label. The process of improvement is graphically described in Figure 4, left.

Candidate labels and confidence values are obtained from *knn*, the annotation method. Semantic association between labels is obtained by measuring co-occurrences of labels at document level in an external

⁶The training set was created for applying a recently proposed semi-supervised learning algorithm based on unlabeled data [13], however time constrains do not allowed us to use such an algorithm.

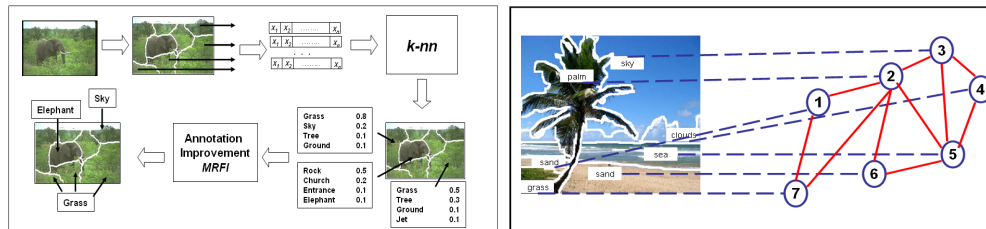


Figure 4: Left: graphical description of the improvement process of *MRFI*. Right: graphical representation of the *MRF*; (red) line-arcs consider semantic cohesion between labels, while (blue) dashed-arcs consider relevance weight of each label according to *knn*.



Figure 5: Expansion of the topic 36 using annotations. Annotations are showed below each segmented image. The expanded query is shown below image's annotations.

corpus of documents. For each region *knn* ranks labels in decreasing order of the relevance of the labels to being the correct annotation for such region. We consider the set of the top- k more relevant labels for each region. In this way we have k -candidate labels for each region in each image, each of these candidate labels is accompanied by a relevance weight, according to the *knn* ranking. Taking into account these relevance weights for the region-label assignments *MRFI* assigns to each region within a given image the label that is more likely, given the labels assigned to neighboring regions in the same image. Intuitively *MRFI* selects the best configuration of regions-labels assignments for each image, given the semantic cohesion between labels assigned to spatially connected regions and given the relevance weight of each region-label assignment. For doing this the set of regions-labels assignments for a given image is modeled with a Markov random field (*MRF*), see Figure 4, right. Each possible assignment of regions-labels for the image is said a configuration of the *MRF*. The goal of *MRFI* is to select the (pseudo) optimal configuration by considering the relevance of each assignment and the semantic association between labels assigned to neighboring regions. The optimal configuration for each image is the configuration that minimizes an energy function defined by potentials. Each potential is a function that considers one aspect of the problem at hand. A first potential function considers the relevance weight of each region-label assignment, while another potential attempts to keep semantic cohesion between labels assigned to connected regions. The minimization of the energy function is carried out by simulated annealing. The best configuration is then considered the correct annotation of each image. For further details about this method we suggest the reader to follow the references [8].

5.3 Query/document expansion

For document expansion the annotations assigned to each region in each image in the *IAPR-TC 12 Benchmark* are added to the original annotation. For query expansion the sample topic images were segmented and annotated, then the set of annotations was used for expanding (or creating) textual queries, in Figure 5 an expanded topic is shown. As we can see some labels are repeated on the resulting expanded topic (*sky, people and tree*); we decided to consider repeated labels because in this way repeated terms are considered representative terms of the expanded query. If we would not considered label's repetitions the word *people*,

Languages	Run-ID	Methods	Type	MAP	Ranking
English-English	<i>INAOE-EN-EN-NaiveQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	22 / 142
Dutch-English*	<i>INAOE-NL-EN-NaiveWBQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	1 / 4
French-English	<i>INAOE-FR-EN-NaiveQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	3 / 21
German-English	<i>INAOE-DE-EN-NaiveQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	3 / 20
Italian-English	<i>INAOE-IT-EN-NaiveWBQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	3 / 10
Japanese-English	<i>INAOE-JA-EN-NaiveWBQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	2 / 6
Portuguese-English	<i>INAOE-PT-EN-NaiveWBQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	2 / 9
Russian-English	<i>INAOE-RU-EN-NaiveWBQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	2 / 6
Spanish-English	<i>INAOE-ES-EN-NaiveWBQE-IMFB</i>	NQE+IMFB	Mixed	0.1986	2 / 9
Visual-English*	<i>INAOE-VISUAL-EN-AN-EXP-3</i>	NQE+ABE+IMFB	Mixed	0.1925	1 / 1
German-German	<i>INAOE-DE-DE-NaiveQE-LATE-FUSION</i>	NQE+LF	Mixed	0.1341	13 / 30
English-German	<i>INAOE-EN-DE-NaiveWBQE-LATE-FUSION</i>	NQE+LF	Mixed	0.1113	11 / 17
Spanish-Spanish	<i>INAOE-ES-ES-NaiveQE-LATE-FUSION</i>	NQE+LF	Mixed	0.1481	5 / 15
English-Spanish	<i>INAOE-EN-ES-NaiveWBQE-LATE-FUSION</i>	NQE+LF	Mixed	0.1145	2 / 6
Dutch-Random*	<i>INAOE-NL-RND-NaiveQE</i>	NQE	Text	0.0828	1 / 2
English-Random	<i>INAOE-EN-RND-NaiveQE-IMFB</i>	NQE+IMFB	Mixed	0.1243	6 / 11
French-Random	<i>INAOE-FR-RND-NaiveQE-IMFB</i>	NQE+IMFB	Mixed	0.1243	3 / 10
German-Random	<i>INAOE-DE-RND-NaiveQE-IMFB</i>	NQE+IMFB	Mixed	0.1243	4 / 11
Italian-Random*	<i>INAOE-IT-RND-NaiveQE</i>	NQE	Text	0.0798	1 / 2
Portuguese-Random*	<i>INAOE-PT-RND-NaiveQE</i>	NQE	Text	0.0296	1 / 2
Russian-Random*	<i>INAOE-RU-RND-NaiveQE</i>	NQE	Text	0.0763	1 / 2
Spanish-Random*	<i>INAOE-ES-RND-NaiveQE-IMFB</i>	NQE+IMFB	Mixed	0.1243	1 / 5

Table 2: Top ranked entries for each of the query-target language configurations comprised in the *TIA*’s submitted runs. Entries marked with an asterisk indicate configurations in which *TIA* was the only participant group.

for example, would be considered equally important (for representing the query) if it appeared as annotation of an unique region in a single image or as 4 annotations of 4 different regions within the three topic’s images.

6 Experimental results

A total of 95 runs were submitted to the *ImageCLEF2007*, comprising all of the above described methods: intermedia feedback (*IMFB*), naive web-based query expansion (*NQE*), repetition web-based query expansion (*WBQE*), late fusion of independent retrievers (*LF*) and annotation based expansion (*ABE*). Note that some runs are combination of these methods. The top ranked entries for each language configuration together with a description of the methods used are shown in Table 2.

As we can see, most of the entries are ranked near the top one. The best performance overall runs were obtained by using *IMFB* together with *NQEs*. We can observe that the runs with *IMFB+NQE* for target language English have exactly the same *MAP* value, independently of the query language. This is due to the fact that all of the queries were translated to English, then the snippets returned by Google^R were used for expanding the translated query. Snippets were similar because the English search engine was used, and translations were very similar for all the query languages. The expanded queries were then combined with the captions of the top–10 relevant documents according to the baseline *FIRE* run, which were the same for all the query languages.

Other entries ranked high were those using *NQE*. Actually the *NQE* is present in all of the top ranked runs. *NQE* outperformed *WBQE* in all of the language configurations. This can be graphically appreciated in Figure 6 (left). This is a surprising result because with the naive approach several noisy terms are added to the queries. While with the repetition approach only the terms that most appear among all the snippets are added. Therefore, the good results of *NQE* are due to the inclusion of many highly related terms, while the insertion of some noisy terms does not affect the performance of the retrieval model. From Figure 6 (left) we can also clearly appreciate that *IMFB* outperformed the *LF* approach.

Six runs based on annotation expansion were submitted to the *ImageCLEF2007*. In these runs document and query expansion was combined with the other techniques proposed in previous sections. The descriptions of the annotation based expansion (*ABE*) runs submitted to *ImageCLEF2007* are shown in Table 3, together with their overall ranking position. Results with *ABE* are mixed. The two top ranked runs with *ABE*

ID	ID	Methods	ABQE	ABDE	Ranking
1	<i>INAOE-EN-EN-AN-EXP-5</i>	Baseline,IMFB	X	-	57
2	<i>INAOE-VISUAL-EN-AN-EXP-3</i>	Baseline,IMFB	X	X	58
3	<i>INAOE-EN-EN-NaiveWBQE-ANN-EXP-LF-2</i>	NQE,LF,	X	X	84
4	<i>INAOE-EN-EN-AN-EXP-2</i>	NQE,Baseline	X	X	133
5	<i>INAOE-EN-EN-DQE-ANN-EXP-LF</i>	NQE,LF	X	-	389
6	<i>INAOE-EN-EN-AN-EXP-1</i>	Baseline	X	X	447

Table 3: Settings of the annotation expansion based runs. An **X** indicates that the corresponding technique is used. *ABQE* is for annotation-based query expansion and *ABDE* is for annotation-based document expansion. The ranking position is shown, a total of 474 runs were submitted for the *IAPR-TC 12 Benchmark 2007* Collection.

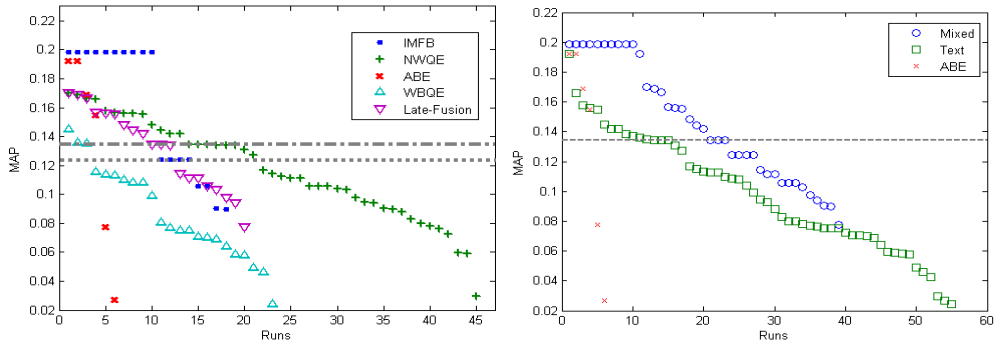


Figure 6: Left: *MAP* value for entries based on intermedia feedback (*IMFB*), late fusion (*LF*), naive Web-based query expansion (*NQE*), repetition web-based query expansion (*WQBE*), and annotation based expansion (*ABE*); the upper dashed line indicates the performance of our English-English baseline; while the dotted line indicates the average *MAP* overall the submitted runs. Right: Comparison of the *MAP* value obtained by mixed and text-only runs. *MAP* of *ABE* entries is shown too. The dashed line indicates the average *MAP* value over all submitted runs.

correspond to entries that used *ABE+IMFB*. This the result can be due to the *IMFB* performance instead of the *ABE* technique. The third *ABE* ranked run used *ABE* of documents and queries with *NQE+LF* which obtained a slightly lower *MAP* than *NQE+LF* without *ABE*. Therefore no gain can be attributed to the *ABE* technique. The other *ABE* runs were ranked low. Summarizing, relatively good results were obtained with *ABE*, though this results can be due to the other techniques employed (*IMFB* and *NQE*). However, we should emphasize that this was our very first effort towards developing annotation based methods for improving image retrieval. In Figure 6 (right) we compare mixed and text-only runs. From this Figure it is clear that the mixed approaches were always superior to the text-only runs. This result clearly illustrates the fact that performance of independent retrievers can be improved by considering both modalities for image retrieval.

7 Conclusions

We have described the first participation of *INAOE's TIA* research group at the *ImageCLEF2007* photographic retrieval task. A total of 95 runs were submitted comprising most of the query languages and all of the target ones. Experiments were performed with two widely used approaches on image retrieval from annotated collections: intermedia feedback and late fusion of independent retrievers. A Web-based query expansion technique was proposed for introducing context into the topics and a novel annotation based expansion technique was also developed. The best results were obtained with the intermedia feedback combined with a version of the web-based query expansion technique. With the naive expansion technique several relevant related terms are added to the original query, as well as many noisy terms. However noisy terms do not affected the performance of the retrieval system. The intermedia feedback technique outperformed late fusion by a large margin.

Relatively good results were obtained with the annotation based runs, however this can be due to the

techniques combined with the *ABE* methods. Results with this technique give evidence that automatic image annotation methods could be helpful on image retrieval from annotated collections. This because promising results were obtained even when segmentation was poor, the training set was extremely small and imbalanced, and annotations did not covered the objects present within the image collection. In consequence for future work we will address all of these issues by creating a larger training set of annotated images, using other segmentation algorithms, defining labels objectively trying to cover most of the objects present in images within the collection and keeping balanced the training set. Also we intent to use other annotation methods.

Acknowledgements. We would like to thank M. Grubinger, A. Harbury, T. Deselaers, P. Clough and all other organizers of *ImageCLEF2007* because of their important effort. This work was partially supported by CONACyT under grant 205834.

References

- [1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Pearson E. L., 1999.
- [2] 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.
- [3] Y. Chang, W. Lin, and Hsin-Hsi Chen. Combining text and image queries at imageclef 2005. In *Working Notes of the CLEF*. CLEF, 2005.
- [4] 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.
- [5] P. Clough, H. Müeller, T. Deselaers, M. Grubinger, T. Lehmann, J. Jensen, and W. Hersh. The clef 2005 cross-language image retrieval track. In *Working Notes of the CLEF 2005*. CLEF, 2005.
- [6] R. Datta, J. Li, and J. Z. Wang. Content-based image retrieval - approaches and trends of the new age. In *Proceedings ACM International Workshop on Multimedia Information Retrieval*, Singapore, 2005. ACM Multimedia.
- [7] H. J. Escalante, M. Montes, and L. E. Sucar. Improving automatic image annotation based on word co-occurrence. In *Proceedings of the 5th International Adaptive Multimedia Retrieval workshop*, Paris, France, July 2007.
- [8] H. J. Escalante, M. Montes, and L. E. Sucar. Word co-occurrence and mrf's for improving automatic image annotation. In *In Proceedings of the 18th British Machine Vision Conference (BMVC 2007) To appear*, Warwick, UK, September, 2007.
- [9] M. Grubinger, P. Clough, A. Hanbury, and H. Müller. Overview of the ImageCLEF 2007 photographic retrieval task. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.
- [10] M. Grubinger, P. Clough, H. Müller, and T. Deselaers. The iapr tc-12 benchmark: A new evaluation resource for visual information systems. 2005.
- [11] J. S. Hare, P. H. Lewis, P. G.B. Enser, and C. J. Sandom. Mind the gap: Another look at the problem of the semantic gap in image retrieval. In Hanjalic A. Chang, E. Y. and Eds. Sebe, N., editors, *Proceedings of Multimedia Content Analysis, Management and Retrieval*, volume 6073, San Jose, California, USA, 2006. SPIE.
- [12] N. Maillot, J. 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.
- [13] H. Marin-Castro, L. E. Sucar, and E. F. Morales. Automatic image annotation using a semi-supervised ensemble of classifiers. In *In Proceedings of the 12th Iberoamerican Congress on Pattern Recognition (CIARP 2007), To appear*, 2007.
- [14] M. M. Rahman, V. Sood, B. C. Desai, and P. Bhattacharya. Cindi at imageclef 2006: Image retrieval and annotation tasks for the general photographic and medical image collections. In *Working Notes of the CLEF*. CLEF, 2006.
- [15] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.
- [16] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [17] D. Zimepekis and E. Gallopoulos. Tmg: A matlab toolbox for generating term-document matrices from text collections. In C. Nicholas J. Kogan and M. Teboulle, editors, *Grouping Multidimensional Data: Recent Advances in Clustering*, pages 187–210. Springer, 2005.