CLEF2007 image annotation task: An SVM-based cue integration approach

CLEF2007 Image Annotation Task: an SVM-based Cue Integration Approach

Tatiana Tommasi, Francesco Orabona, and Barbara Caputo
IDIAP Research Institute,
Centre Du Parc, Av. des Pres-Beudin 20,
P. O. Box 592, CH-1920 Martigny, Switzerland
{ttommasi, forabona, bcaputo}@idiap.ch

Abstract

This paper presents the algorithms and results of our participation to the medical image annotation task of ImageCLEFmed 2007. We proposed, as a general strategy, a multi-cue approach where images are represented both by global and local descriptors, so to capture different types of information. These cues are combined during the classification step following two alternative SVM-based strategies. The first algorithm, called Discriminative Accumulation Scheme (DAS), trains an SVM for each feature type, and considers as output of each classifier the distance from the separating hyperplane. The final decision is taken on a linear combination of these distances: in this way cues are accumulated, thus even when they both are misleaded the final result can be correct. The second algorithm uses a new Mercer kernel that can accept as input different feature types while keeping them separated. In this way, cues are selected and weighted, for each class, in a statistically optimal fashion. We call this approach Multi Cue Kernel (MCK). We submitted several runs, testing the performance of the single-cue SVM and of the two cue integration methods. Our team was called BLOOM (BLanceflOr-tOMed.im2) from the name of our sponsors. The DAS algorithm obtained a score of 29.9, which ranked fifth among all submissions. We submitted two versions of the MCK algorithm, one using the one-vs-all multiclass extension of SVMs and the other using the one-vs-one extension. They scored respectively 26.85 and 27.54, ranking first and second among all submissions.

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**]: Database Applications—*Image databases*; I.5 [**Pattern Recognition**]: I.5.2 Design Methodology

General Terms

Measurement, Performance, Experimentation

Keywords

Automatic Image Annotation, Cue Integration, Support Vector Machines, Kernel Methods

1 Introduction

The amount of medical image data produced nowadays is constantly growing, with average-sized radiology departments producing several tera-bites of data annually. The cost of manually annotating these images is very high; furthermore, manual classification induces errors in the tag assignment, which means that a part of the available knowledge is not accessible anymore to physicians [5]. This calls for automatic annotation algorithms able to perform the task reliably, and benchmark evaluations are thus extremely useful for boosting advances in the field. The ImageCLEFmed annotation task has been established in 2005, and in 2007 it has provided participants with 11000 training and development data, spread across 116 classes. The task consisted in assigning the correct label to 1000 test images. For further informations on the annotation task of ImageCLEF 2007 we refer the reader to [6].

This paper describes the algorithms submitted by the BLOOM (BLanceflOr-tOMed.im2) team, at its first participation to the CLEF benchmark competition. In order to achieve robustness, a crucial property for a reliable automatic system, we opted for a multi-cue approach, using raw pixels as global descriptors and SIFT features as local descriptors. The two feature types were combined together using two different SVM-based integration schemes. The first is the Discriminative Accumulation Scheme (DAS), proposed first in [7]. For each feature type, an SVM is trained and its output consists of the distance from the separating hyperplane. Then, the decision function is built as a linear combination of the distances, with weighting coefficients determined via cross validation. We submitted a run using this method (BLOOM-BLOOM_DAS) that obtained a score of 29.9, ranking fifth among all submissions.

The second integration scheme consists in designing a new Mercer kernel, able to take as input different feature types for each image data. We call it Multi Cue Kernel (MCK); the main advantage of this approach is that features are selected and weighted during the SVM training, thus the final solution is optimal as it minimizes the structural risk. We submitted two runs using this algorithm, the first (BLOOM-BLOOM_MCK_oa) using the one-vs-all multiclass extension of SVM; the second (BLOOM-BLOOM_MCK_oo) using instead the one-vs-one extension. These two runs ranked first and second among all submissions, with a score of respectively 26.85 and 27.54. These results overall confirm the effectiveness of using multiple cues for automatic image annotation.

The rest of the paper is organized as follows: section 2 describes the two types of feature descriptors we used at the single cue stage. Section 3 gives details on the two alternative SVM-based cue integration approaches. Section 3 reports the experimental procedure adopted and the results obtained, with a detailed discussion on the performance of each algorithm. The paper concludes with a summary discussion.

2 Single Cue Image Annotation

The aim of the automatic image annotation task is to classify images into a set of classes. In particular classes were defined in relation to the four independent axis of modality, body orientation, body region, and biological system examined, according to the IRMA code [3]. The labels are hierarchical therefore, errors in the annotation are counted depending on the level at which the error is done and on the number of possible choices. For each image the error ranges from 0 to 1, respectively if the image is correctly classified or if the predicted label is completely wrong. It is also possible to assign a "don't know" label, in this case the score is 0.5.

The strategy we propose is to extract a set of features from each image and to use then a Support Vector Machine (SVM) to classify the images. We have explored a local approach, using SIFT descriptors, and a global approach, using the raw pixels.

2.1 Feature Extraction

We explored the idea of "bag of words" for classification, a common concept in many state of the art approaches in images classification. This is based on the idea that it is possible to transform the images into a set of prespecified visual words, and to classify the images using the statistics of appearance of each word as feature vectors.

Most of these systems are based on the use of the SIFT descriptor [4]. The basic idea of SIFT is to describe an area of an image in a way that is robust to noise, illumination, scale, translation and rotation changes. The SIFT points are selected in the image as local maxima of the scale-space, in this sense the SIFT points are intrinsically easy to be tracked. Despite the usefulness of SIFT, there is no reason to believe that these points are the most informative for a classification task. This has been pointed out by different works and systematically verified by [8]. In that work it is shown that a dense random sampling of the SIFT points is always superior to any strategy based on interest points detectors. Moreover due to the low contrast of the radiographs it would be difficult to use any interest point detector. So in our approach we densely sampled each input image, extracting in each point a SIFT descriptor.

Another modification we made is based on the fact that the rotation invariance could be useless for the ImageCLEF classification task, as the various structures present in the radiographs are likely to appear always with the same orientation. Moreover the scale is not likely to change too much between images of the same class, so we extracted the SIFT at only one octave, the one that gave us the best classification performances. In this sense we have decoupled the extraction of a SIFT keypoint from the description of the point itself. To keep the complexity of the description of each image low and at the same to retain as much information as possible, we matched each extracted SIFT with a number of template SIFTs. These template SIFTs form our vocabulary of visual words. It is built using a standard K-means algorithm, with K equal to 500, on a random collection of SIFTs extracted from the training images. Various sizes of vocabulary were tested with no significant differences, so we have chosen the smaller one with good recognition performances. Note that in this phase also testing images can be used, because the process is not using the labels and it is unsupervised. At this point each image could be described with the raw counts of each visual word.

To add some kind of spatial information to our features we divided the images in four subimages, collecting the histograms separately for each subimage. In this way the dimension of the input space is multiplied by four, but in our tests we gained about 3% of classification performances. We have extracted 1500 SIFT in each subimage: such dense sampling adds robustness to the histograms. See Figures 1 and 2 for an example.

Another approach that we explored was the simplest possible global description method: the raw pixels. The images were resized to 32x32 pixels, regardless of the original dimension, and normalized to have sum equal to one, then the 1024 raw pixels values were used as input features. This approach is at the same time a baseline for the classification system and a useful "companion" method to boost the performance of the SIFT based classifier (see section 2.2).

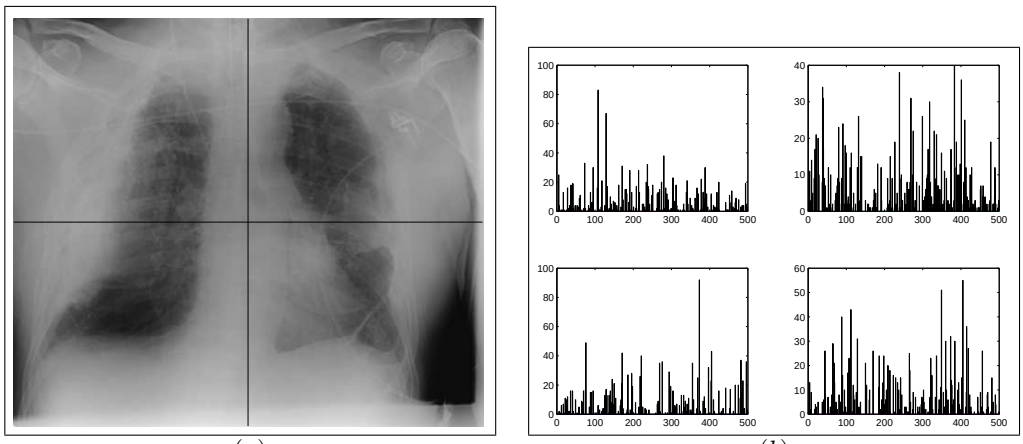
2.2 Classification

For the classification step we used an SVM with an exponential χ^2 as kernel, for both the local and global approaches:

$$K(X,Y) = \exp\left(-\gamma \sum_{i=1}^{N} \frac{(X_i - Y_i)^2}{X_i + Y_i}\right). \quad (1)$$

The parameter γ was tuned through cross-validation (see section 4). This kernel has been successfully applied for histogram comparison and it has been demonstrated to be positive definite [2], thus it is a valid kernel.

Even if the labels are hierarchical, we have chosen to use the standard multi-class approaches. This choice is motivated by the finding that, with our features, the recognition rate was lower using an axis-wise classification. This could be due to the fact that each super-class has a variability so



(a) (b)

Figure 1: (a) Radiographic image divided in 4 subimages and (b) corresponding counts of the visual words.

high that the chosen features are not able to model it, while they can very well model the small sub-classes. In particular we have tested both one-vs-one and one-vs-all multi-class extension for SVM.

3 Multi Cue Annotation

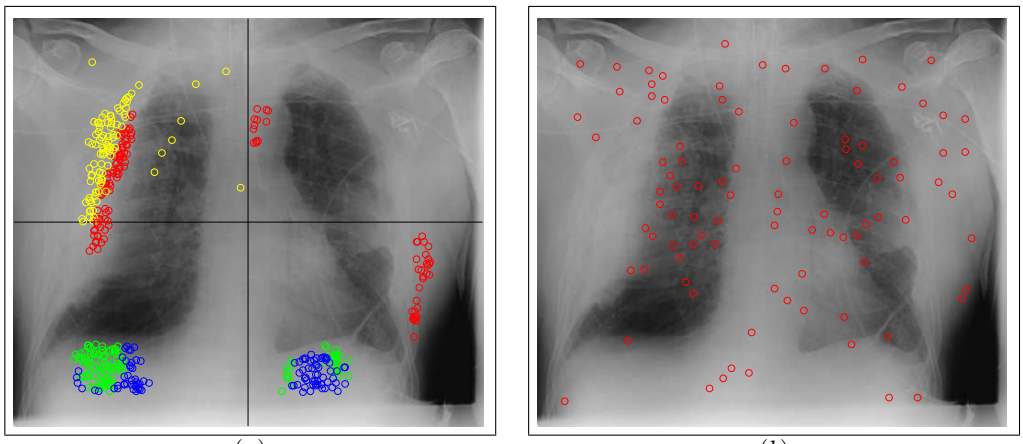
Due to the fundamental difference in how local and global features are computed it is reasonable to suppose that the two representations provide different kinds of information. Thus, we expect that by combining them through an integration scheme, we should achieve a better performance, namely higher classification performance and higher robustness.

In the computer vision and pattern recognition literature some authors have suggested different methods to combine information derived from different cues (for a review on the topic we refer the reader to [9]). Some of them are based on building new representations, but this technique does not solve the robustness problem because if one of the cues gives misleading information it is quite probable that the new feature vector will be adversely affected. Moreover, the dimension of such a feature vector would increase as the number of of cues grows, implying longer learning and recognition times, greater memory requirements and possibly curse of dimensionality effects. The strategy we follow in this paper is to use integration schemes, thus keeping the feature descriptors separated and fusing them at a mid- or high- level. In the rest of the section we describe the two alternative integration schemes we used in the ImageCLEF competition. The first, the Discriminative Accumulation Scheme (DAS, [7]), is a high-level integration scheme, meaning that each single cue first generate a set of hypotheses on the correct label of the test image, and then those hypotheses are combined together so to obtain a final output. This method is described in section 3.1. The second, the Multi Cue Kernel (MCK), is a mid-level integration scheme, meaning that the different features descriptors are kept separated but they are combined in a single classifier generating the final hypothesis. This algorithm is described in section 3.2.

3.1 Discriminative Accumulation Scheme

The Discriminative Accumulation Scheme is an integration scheme for multiple cues that does not neglect any cue contribution. It is based on a weak coupling method called accumulation. The main idea of this method is that information from different cues can be summed together.

Suppose we are given M object classes and for each class, a set of N_j training images $\{I_i^j\}_{i=1}^{N_j}$, $j = 1, \dots M$. For each image, we extract a set of P different cues:



(a) (b)

Figure 2: Difference between random sampling and interest point detector. In (a) the four most present visual words in the image are drawn, each with a different color. In (b) the result of standard SIFT extraction in the same octave used in (a).

$$T_p = T_p(I_i^j), \quad p = 1 \dots P \tag{2}$$

so that for an object j we have P new training sets $\{T_p(I_i^j)\}_{i=1}^{N_j}, j = 1, \dots M, p = 1 \dots P$. For each we train an SVM. Kernel functions may differ from cue to cue and model parameters can be estimated during the training step via cross validation. Given a test image $\hat{I}$ and assuming $M \geq 2$, for each single-cue SVM we compute the distance from the separating hyperplane:

$$D_j(p) = \sum_{i=1}^{m_j^p} \alpha_{ij}^p y_{ij} K_p \left(T_p(I_i^j), T_p(\hat{I}) \right) + b_j^p. \tag{3}$$

After collecting all the distances $\{D_j(p)\}_{p=1}^P$ for all the j objects $j = 1, \dots, M$ and the p cues $p = 1, \dots, P$, we classify the image $\hat{I}$ using the linear combination:

$$j^* = \underset{j=1}{\overset{M}{\operatorname{argmax}}} \{ \sum_{p=1}^{P} a_p D_{j(p)} \}, \quad a_p \in \Re^+. \tag{4}$$

The coefficients $\{a_p\}_{p=1}^P$ are evaluated via cross validation during the training step.

3.2 Multi Cue Kernel

DAS can be defined a high-level integration scheme, as fusion is performed as a post-processing step after the single-cue classification stage. As an alternative, we developed a mid-level integrating scheme based on multi-class SVM with a Multi Cue Kernel K_{MC}. This new kernel combines different features extracted form images; it is a Mercer kernel, as positively weighted linear combination of Mercer kernels are Mercer kernels themselves [1]:

$$K_{MC}(\{T_p(I_i)\}_p, \{T_p(I)\}_p) = \sum_{p=1}^{P} a_p K_p(T_p(I_i), T_p(I)). \tag{5}$$

In this way it is possible to perform only one classification step, identifying the best weighting factors a_p while optimizing the other kernel parameters. Another advantage of this approach is that it makes it possible to work both with one-vs-all and one-vs-one SVM extensions to the multiclass problem.

4 Experiments

Our experiments started evaluating the performance of local and global features separately before testing our integration methods. Two sets of experiments using single-cue SVM were ran to select the best kernel parameters through cross validation. The original dataset was divided in three parts: training, validation and testing. We merged them together and extracted 5 random and disjoint train/test splits of 10000/1000 images. We considered as the best parameters the one giving the best average score on the 5 splits. Note that, according to the method used for the score evaluation, the best average score is not necessary the best recognition rate. Besides obtaining the optimal parameters, these experiments showed that the SIFT features outperform the raw pixel ones. It could be predictable since the last year ImageCLEF competition results showed that local features are generally more informative than global features for the annotation task.

Then we adopted the same experimental setup for DAS and MCK. In particular for DAS we used the distances from the separating hyperplanes associated with the best results of the previous step, so the cross validation was used only to search the best weights for cue integration. On the other hand, for MCK the cross validation was applied to look for the best kernel parameters and the best feature's weights at the same time. In both cases weights could vary from 0 to 1.

Finally we used the results of the previous phases to run our submission experiment on the 1000 unlabeled images of the challenge test set using all the 11000 images of the original dataset as training.

The ranking, name and score of our submitted runs together with the score gain respect to the best run of other participants are listed in Table 1. Our two runs based on the MCK algorithm ranked first and second among all submissions stating the effectiveness of using multiple cues for automatic image annotation. It is interesting to note that even if DAS has a higher recognition rate, its score is worse than that obtained using the feature SIFT alone. This could be due to the fact that when the label predicted by the global approach, the raw pixels, is wrong, the true label is far from the top of the decision ranking.

In Table 2 there is a summary of the parameters used for our runs and the number of support vectors obtained. As we could expect, the best feature weight (see (4) and (5)) for SIFT results higher than that for raw pixels for all the integration methods. The number of support vectors for the MCK run using one-vs-one multiclass SVM extension (MCK_oa) is slightly higher than that used by the single cue SIFT_oa but lower than that used by PIXEL_oa. For the MCK run using one-vs-one multiclass SVM extension (MCK_oo) the number of support vectors is even lower than that of both the single cues SIFT_oo and PIXEL_oo. These results show that combining two features with the MCK algorithm can simplify the classification problem. For DAS we counted the support vectors summing the ones from SIFT_oa and PIXEL_oa but considering only once the support vectors associated with the training images that resulted in common between the single cues. The number of support vector for DAS exceed that obtained for both MCK_oa and MCK_oo showing a higher complexity of the classification problem.

Table 3 shows in details some examples of classification results. The first, second and third column contain examples of images misclassified by one of the two cues but correctly classified by DAS and MCK_oa. The fourth column shows an example of an image misclassified by both cues and by DAS but correctly classified by MCK_oa. It is interesting to note that combining local and global features can be useful to recognize images even if they are compromised by the presence of artifacts that for medical images can be prosthesis or reference labels put on the acquisition screen.

A deeper analysis of our results can be done considering the performance of the single-cue, discriminative accumulation and multicue kernel approach for each class. In Table 4 the number

Rank	Name	Score	Gain	Rec. rate
1	BLOOM-BLOOM_MCK_oa	26.8470167911	4.0828086669	89.7%
2	BLOOM-BLOOM_MCK_oo	27.5449911826	3.3848342754	89.0%
3	BLOOM-BLOOM_SIFT_oo	28.7301320009	2.1996934571	88.4%
4	BLOOM-BLOOM_SIFT_oa	29.45575794	1.474067518	88.5%
5	BLOOM-BLOOM_DAS	29.9033537771	1.0264716809	88.9%
28	BLOOM-BLOOM_PIXEL_oa	68.2130545639	−37.2832291059	79.9%
29	BLOOM-BLOOM_PIXEL_oo	72.410704904	−41.4808794460	79.2%

Table 1: Ranking of our submitted runs, name, score and gain respect to the best run of the other participants.

Rank	Name	γ_{sift}	γ_{pixel}	C	a_{sift}	a_{pixel}	#SV
1	BLOOM-BLOOM_MCK_oa	0.5	5	5	0.80	0.20	7916
2	BLOOM-BLOOM_MCK_oo	0.1	1.5	20	0.90	0.10	7037
3	BLOOM-BLOOM_SIFT_oo	0.05		40			7173
4	BLOOM-BLOOM_SIFT_oa	0.25		10			7704
5	BLOOM-BLOOM_DAS	0.25	5	10	0.76	0.24	9090
28	BLOOM-BLOOM_PIXEL_oa		5	10			8329
29	BLOOM-BLOOM_PIXEL_oo		3	20			7381

Table 2: Here are shown the best parameters obtained by cross validation and used for the classification, together with the number of Support Vectors for each of our submitted runs.

of images correctly recognized for each class are listed and it is possible to note that in few cases PIXEL_oa outperforms SIFT_oa, and to observe where MCK_oa outperforms both SIFT_oa and DAS. The difference between our approaches can be better evaluated considering the confusion matrices. They are shown as images in Figure 3. We ordered the classes following the way in which they are listed in table 4 and used a colormap corresponding to the number of images varying from zero to five to let the misclassified images stand out. It is clear that our methods differ principally for how the wrong images are labeled. The more matrices present sparse values out of the diagonal and far away from it, the worse the method is.

				
PIXEL_oa	11°	**1°**	12°	5°
SIFT_oa	**1°**	2°	2°	5°
DAS	**1°**	**1°**	**1°**	2°
MCK_oa	**1°**	**1°**	**1°**	**1°**

Table 3: Example of images misclassified by one or both cues and correctly classified by DAS or MCK.The values correspond to the decision rank.

CLASS	MCK_oa	SIFT_oa	DAS	PIXEL_oa	TOT	CLASS	MCK_oa	SIFT_oa	DAS	PIXEL_oa	TOT	CLASS	MCK_oa	SIFT_oa	DAS	PIXEL_oa	TOT	CLASS	MCK_oa	SIFT_oa	DAS	PIXEL_oa	TOT
1121-110-213-700	0	0	0	0	3	**1121-120-516-700**	2	2	2	2	4	**1121-210-331-700**	0	0	0	0	1	**1121-240-441-700**	4	4	4	4	5
1121-110-411-700	13	11	12	8	14	**1121-120-517-700**	2	2	2	3	3	**1121-220-213-700**	2	2	2	1	2	**1121-240-442-700**	4	4	4	4	4
1121-110-414-700	38	38	38	35	38	**1121-120-800-700**	22	22	22	22	22	**1121-220-230-700**	17	17	17	17	17	**1121-320-941-700**	10	10	10	10	10
1121-110-415-700	9	9	8	6	9	**1121-120-911-700**	5	5	5	5	6	**1121-220-310-700**	7	7	7	6	7	**1121-420-212-700**	4	2	3	4	4
1121-115-700-400	13	13	13	12	13	**1121-120-914-700**	6	6	6	6	6	**1121-220-330-700**	0	0	0	0	1	**1121-420-213-700**	4	4	4	4	4
1121-115-710-400	2	2	2	2	3	**1121-120-915-700**	5	5	5	5	6	**1121-228-310-700**	1	1	1	0	1	**1121-430-213-700**	8	8	8	6	9
1121-116-917-700	1	1	1	1	1	**1121-120-918-700**	2	2	2	3	3	**1121-229-310-700**	1	0	1	1	1	**1121-430-215-700**	0	0	0	0	1
1121-120-200-700	33	33	33	33	34	**1121-120-919-700**	2	2	2	1	2	**1121-230-462-700**	2	2	2	2	2	**1121-460-216-700**	1	1	1	2	2
1121-120-310-700	20	20	20	20	20	**1121-120-921-700**	9	9	9	5	9	**1121-230-463-700**	5	5	5	5	5	**1121-490-310-700**	1	1	1	0	1
1121-120-311-700	3	3	3	3	3	**1121-120-922-700**	11	11	11	7	11	**1121-230-911-700**	0	0	0	0	2	**1121-490-415-700**	6	6	6	6	6
1121-120-320-700	11	11	11	9	11	**1121-120-930-700**	0	0	0	0	2	**1121-230-914-700**	1	1	1	0	1	**1121-490-915-700**	4	4	4	4	4
1121-120-330-700	22	23	22	20	23	**1121-120-933-700**	0	0	0	0	1	**1121-230-915-700**	1	1	1	1	1	**1122-220-333-700**	0	0	0	0	0
1121-120-331-700	1	0	1	1	1	**1121-120-934-700**	2	1	2	1	2	**1121-230-921-700**	7	7	7	6	7	**1123-110-500-000**	84	78	80	69	91
1121-120-413-700	3	3	3	0	3	**1121-120-942-700**	9	10	9	7	10	**1121-230-922-700**	6	6	6	6	8	**1123-112-500-000**	0	0	0	0	5
1121-120-421-700	4	4	4	4	5	**1121-120-943-700**	9	9	9	9	10	**1121-230-930-700**	0	0	0	0	1	**1123-121-500-000**	5	5	5	5	8
1121-120-422-700	4	4	4	3	4	**1121-120-950-700**	0	1	0	0	1	**1121-230-934-700**	1	1	1	0	2	**1123-127-500-000**	182	184	184	172	196
1121-120-433-700	1	2	1	1	2	**1121-120-951-700**	1	2	2	0	3	**1121-230-942-700**	9	9	9	9	9	**1123-211-500-000**	89	89	89	88	89
1121-120-434-700	2	2	2	0	2	**1121-120-956-700**	1	0	0	0	2	**1121-230-943-700**	7	6	6	6	7	**1124-310-610-625**	6	6	6	6	6
1121-120-437-700	0	0	0	0	1	**1121-120-961-700**	3	3	3	3	4	**1121-230-950-700**	0	0	0	0	1	**1124-310-620-625**	7	6	6	6	7
1121-120-438-700	0	0	0	0	1	**1121-120-962-700**	5	4	5	3	5	**1121-230-953-700**	0	0	0	0	1	**1124-410-610-625**	7	7	7	7	7
1121-120-441-700	4	4	4	2	5	**1121-127-700-400**	0	0	0	0	3	**1121-230-961-700**	4	4	4	3	4	**1124-410-620-625**	7	7	7	7	7
1121-120-442-700	3	3	3	3	4	**1121-127-700-500**	0	0	0	0	1	**1121-230-962-700**	2	2	2	0	3	**1121-120-91a-700**	0	0	0	0	1
1121-120-451-700	1	1	1	0	1	**1121-129-700-400**	1	1	1	1	1	**1121-240-413-700**	0	0	0	0	2	**1121-12f-466-700**	0	0	0	0	1
1121-120-452-700	0	0	0	0	1	**1121-200-411-700**	9	7	7	4	13	**1121-240-421-700**	4	4	4	3	5	**1121-12f-467-700**	2	2	2	2	2
1121-120-454-700	0	0	0	0	1	**1121-210-213-700**	1	1	1	1	1	**1121-240-422-700**	2	2	2	1	3	**1121-4a0-310-700**	2	2	2	0	2
1121-120-462-700	4	4	4	5	5	**1121-210-230-700**	13	13	13	11	13	**1121-240-433-700**	1	1	1	0	3	**1121-4a0-414-700**	8	7	8	8	8
1121-120-463-700	7	7	7	6	7	**1121-210-310-700**	10	10	10	9	10	**1121-240-434-700**	1	1	1	0	3	**1121-4a0-914-700**	3	3	3	2	5
1121-120-514-700	1	1	1	0	2	**1121-210-320-700**	11	11	11	10	12	**1121-240-437-700**	0	0	0	0	2	**1121-4a0-918-700**	0	1	1	0	1
1121-120-515-700	3	3	3	3	3	**1121-210-330-700**	20	20	20	18	21	**1121-240-438-700**	0	0	0	0	1	**1121-4b0-233-700**	4	4	4	3	4

Table 4: Performance of the single-cue, discriminative accumulation and multicue kernel approach for each class.

5 Conclusions

This paper presented a discriminative multi-cue approach to medical image annotation. We combined global and local information using two alternative fusion strategies, the discriminative accumulation scheme [7] and the multi cue kernel. This last method gave the best performance obtaining a score of 26.85, which ranked first among all submissions.

This work can be extended in many ways. First, we would like to use various types of local and global descriptors, so to select the best features for the task. Second, we would like to add shape descriptors in our fusion scheme, which should result in a better performance. Finally, our algorithm does not exploit at the moment the natural hierarchical structure of the data, but we believe that this information is crucial for achieving significant improvements in performance. Future work will explore these directions.

Acknowledgments

This work was supported by the ToMed.IM2 project (B. C. and F. O), under the umbrella of the Swiss National Center of Competence in Research (NCCR) on Interactive Multimodal Information Management (IM2, www.im2.ch), and by the Blanceflor Boncompagni Ludovisi foundation (T. T., www.blanceflor.se). The support is gratefully acknowledged.

References

[1] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines (and Other Kernel-Based Learning Methods)*. CUP, 2000.

[2] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214–225, 2004.

[3] Schubert Henning Keysers Daniel Kohnen Michael Wein Berthold B. Lehmann, Thomas M. The irma code for unique classification of medical images. In *Proceedings of SPIE Medical Imaging*, volume 5033, pages 440–451, May 2003.

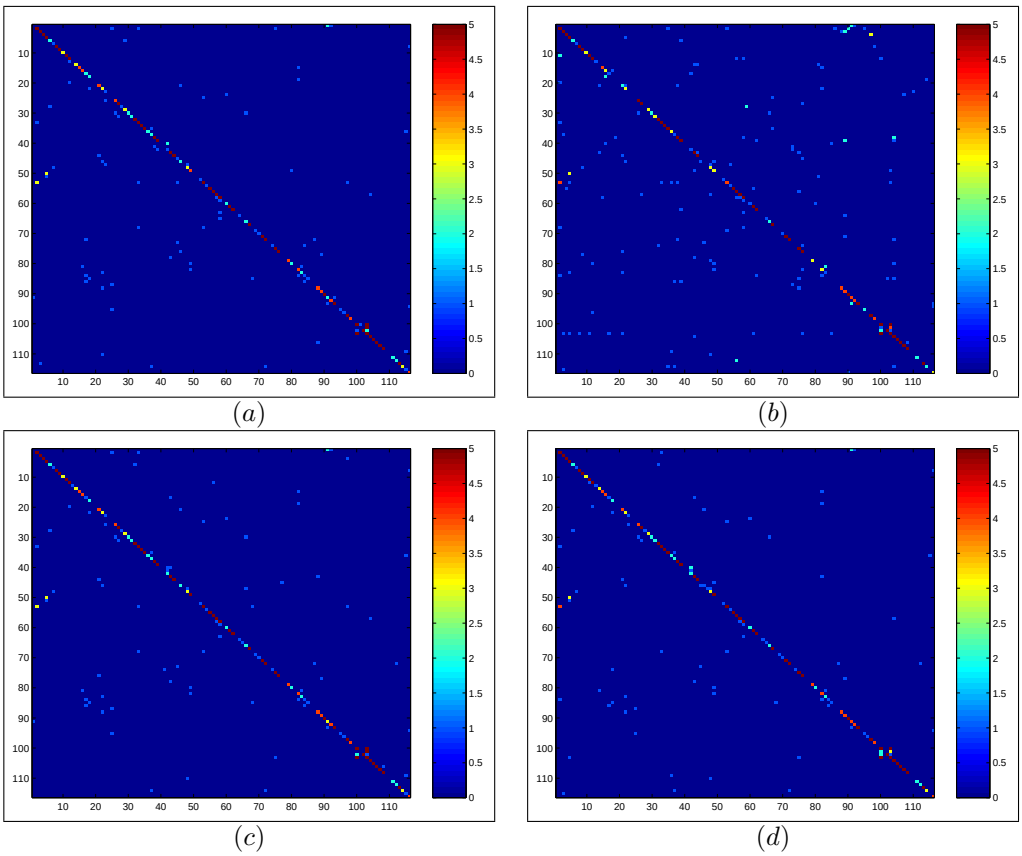


Figure 3: These images represent the confusion matrices respectively for (a) SIFT_oa, (b) Pixel_oa, (c) DAS and (d) MCK_oa. We ordered the classes following the way in which they are listed in table 4 and used a colormap corresponding to the number of images varying from zero to five to let the misclassified images stand out. All the position in the matrices containing five or more images appear dark red.

[4] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157, Washington, DC, USA, 1999. IEEE Computer Society.

[5] M-O-Gueld, M. Kohnen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, and T. M. Lehmann. Quality of dicom header information for image categorization. In *Proceedings of SPIE Medical Imaging*, volume 4685, pages 280–287, 2002.

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

[7] M.E Nilsback and B. Caputo. Cue integration through discriminative accumulation. In *Proceedings of the International conference on Computer Vision and Pattern Recognition*, 2004.

[8] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *Proceedings of the European Conference on Computer Vision*, 2006.

[9] R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45, 2006.