

Fast filtering techniques in medical image classification and retrieval

Xin Zhou, Miaofei Han, Yanli Song, Qiang Li

Medical Image Information Laboratory, Advanced Medical Equipment Research Center, Shanghai Advanced Research Institute, Chinese Academy of Sciences, 99 Haik Road, Building No. 3, Pudong, Shanghai 201210, China
liqiang@sari.ac.cn

Abstract. This article presents the participation of the MIILab (Medical Image Information Laboratory) group in ImageCLEFmed2013. There are three types of tasks for ImageCLEFmed2013: modality classification, image retrieval and compound-image separation. Image modality classification and medical image retrieval are targeted according to MIILab's research interest. The main goal is to perform a feasibility test on applying existing techniques on new applications, such as applying image denoising techniques on image retrieval and classification.

Both global features and local features were employed. Fast filtering techniques were used to obtain global features on color, shape and texture. These global features serves to perform a pre-classification on images. Both low-level and high-level local features were extracted. Bags of features model was used to build final feature vector. Both kNN and SVM classifiers were tried out in modality classification task. Reciprocal kNN was used to perform result fusion in image retrieval task.

The modality classification task was decomposed into a compound image classification and a non-compound image classification. Our approaches achieved 89.9% classification accuracy on training data and 85.1% on testing data for compound image classification. For non compound image classification, accuracy is around 68.3% on training data and 67.7% on testing data. The overall classification accuracy is around 65% on 31 classes. False alarms are mainly from large classes such as compound images (312), X-ray images (101) and organ photos (63), but accuracy per class shows that performance bottleneck also comes from small/medium classes with large content diversity such as statistic figures, chemical structure and 3D images. Best result was around 80% (IBM research lab). For the image retrieval task, one baseline using SURFContext+BoF was submitted and the corresponding MAP (mean average precision) is 0.0086. Even best visual retrieval run obtained only a MAP of 0.018, which is still not comparable with textual approaches (average score 0.2).

1 Introduction

The Cross Language Evaluation Forum (CLEF) organizes contests for the evaluation of information retrieval systems each year. The image retrieval track of

CLEF is called ImageCLEF¹. ImageCLEFmed has been part of ImageCLEF focusing on medical images [1] since 2004. More about the ImageCLEF tasks and results in 2013 can be found in [2, 3].

MIIlab (Medical Image Information Laboratory) is a medical imaging group with research focus in medical imaging technology, particularly those tightly related to advanced medical imaging equipments. This is the first year that MIIlab participates in the medical imaging tasks of ImageCLEF (ImageCLEFmed). The main objective is to perform a feasibility test on applying existing techniques on new applications, such as applying image denoising techniques on image retrieval and classification.

Two tasks of ImageCLEFmed2013 fit MIIlab’s research interests, i.e.: image modality classification and medical image retrieval. The fundamental difficulty for ImageCLEFmed tasks is the diversity of image content. It is due to the fact that the collection contains not only biomedical images, but also figures, mathematic formulas, tables, non-clinical photos, etc. Such a diversity of content generate numerous types of similarity which are difficult to cover in feature space. In this paper, we propose to use fast filtering techniques (monochrome filter, Laplace filter, and line/dot filter [4]) and BoBB(bags of bounding box) to address this problem.

2 Methods

This section describes the basic techniques and collections used by MIIlab in ImageCLEFmed2013. All runs submitted to ImageCLEFmed2013 are purely visual-based. Techniques used can be divided into: 1) pre-processing, 2) features, 3) classifiers and 4) fusion strategies.

2.1 Techniques used

Pre-processing Images in ImageCLEFmed2013 collections are from journal papers, which were mainly converted into JPEG (Joint Photographic Experts Group) format. Certain monochrome images became color images after the conversion, others changed their dynamic range. Hence, images are no longer comparable between them, which introduces additional complication to the task. Normalization on images is required before any other operations take place, and all images are converted into a unified range.

Another issue is the color space. By default JPEG use the RGB (red, green, blue) color space. However, the RGB color spaces is well known for not being tightly corresponded to human color perception and thus less used in image retrieval and classification [5]. On the other hand, it has also been proved that the HSV (Hue, Saturation, Value) color space can give promising results in image retrieval [6, 7].

¹ <http://www.imageclef.org/>

Therefore, the pre-processing step consists of 2 parts. The HSV (Hue, Saturation, Value) color space decomposition and a 0–255 normalization on each channel.

Features In this paper, both global features and local features were employed. Fast filtering techniques, such as monochrome filter, line/dot filter [4], were applied to obtain global features on color, shape and texture. First to third statistical moments were also extracted in certain subregions as low-level local features. SURFContext [8] (an early fusion version of SURF [9] and Shape Context [10]) were implemented to obtain high-level local features. The number of visual keywords is set to be 3000. The classical BoF (bags of features) approach was used to form the final feature vector.

Classifiers Classifiers were only used in modality classification task. Both kNN (k-Nearest Neighbors) and SVM (Support Vector Machine) classifiers were employed, but the latter significantly outperformed the former. Therefore, only SVM classifier-based runs were submitted for evaluation.

Two-level classification is performed: compound image classification and modality classification. The former uses only global features and low-level local features (first to third statistical moments, surface, ratio) of all detectable bounding box. The bounding box detection is based on open source toolbox and the final feature space is built with BoBB. The latter is based on high-level local features with classical BoF approach. Open source machine learning libraries (weka, libsvm) were used with a grid search for parameter selection.

Fusion Techniques Fusion techniques were used in image retrieval task to combine result lists of different techniques. Classical fusion strategies, such as combSUM, combMNZ, combSUM(2)MAX were used with a score normalization based on rank-number based logarithmic weighting function [11]. Recently, a reciprocal kNN based fusion technique was reported to obtain promising result in nature image classification and retrieval [12]. One of the advantage of using this technique is that no score normalization is required. This technique is implemented and labeled as combRKNN in this paper.

2.2 Image Collections

306'538 medical images were available for ImageCLEFmed 2013. Among them, 2'905 images were labeled with modality information and were provided as training data, and 2'582 images were provided without modality information as test data for modality classification task. Details about the setup and collections of the ImageCLEFmed tasks can be found in an overview paper [2].

3 Results

This section describes our results for the two medical tasks.

3.1 Modality Classification

For modality classification task, evaluation is based on the percentage of correctly classified images. For runs of various technique approaches (textual, visual, mixed), the best accuracy and average accuracy together with our results are shown in Table 1.

Table 1. Results of the runs for the modality classification task.

Run ID	best accuracy in %	average accuracy in %
mixed run	81.68	61.97
textual run	64.17	50.79
visual run	80.79	61.24
sari_modality_baseline	66.46	
sari_modality_CCTBB_DRxxDict	65.60	

The average accuracy per class is shown in Table 2. This year we separate compound image classification as an independent task. Once an image was not classified as compound image, it was passed to a non-compound classification for the rest 30 classes.

The overall classification accuracy is around 65% on 31 classes. False alarms are mainly from large classes, such as compound images (COMP), X-ray images (DRXR), etc. Actually, the accuracy in % of COMP and DRXR are around 68%, which limited the overall performance. Certain other classes obtained scores below overall accuracy and became also performance bottleneck. These classes usually contains image of large diversity, such as organ photos (DVOR), non-clinical photos (GNCP), statistic figures (GFIG), tables and forms (GTAB), and 3D images (D3DR). Previous experience shows that compound image classification is of key importance for modality classification. Two specificities exist for compound image classification: 1) compound image share the same image content with 30 other image classes; 2) about 40% of the whole collection are compound images. Hence, a big percentage of incorrectly classified images come from the class of compound images.

Through experiments, it is found that high-level semantic features can hardly provide useful information for compound image classification. However, visual perception can easily make a distinction, as a compound image is consisted of several unconnected figures, invisible bounding boxes are produced in human brain for each figure. Based on connected region detection toolbox in Matlab, the 10 biggest bounding boxes without overlapping were extracted from the value channel of image, and 5 biggest bounding boxes without overlapping were extracted from hue and saturation channels of image. The whole image was always considered as a bounding box. For each bounding box, height, width, ratio, surface, first to third statistical moments (mean, variance, skewness) were used to build a descriptor. Hence, 7 features per bounding box were extracted

Table 2. Results of the runs on each class.

Run id class	number of images	sari_modality_baseline		sari_modality_CCTBB_DRxxDict	
		false alarm	accuracy in %	false alarm	accuracy in %
COMP	1014	312	69.2	320	68.4
D3DR	26	16	38.5	17	34.6
DMEL	20	16	20.0	17	15.0
DMFL	33	7	78.8	9	72.7
DMLI	121	23	81.0	28	76.9
DMTR	20	8	60.0	5	75.0
DRAN	18	5	72.2	5	72.2
DRCO	1	1	0.00	1	0.00
DRCT	186	25	86.6	28	84.9
DRMR	90	27	70.0	26	71.1
DRPE	3	2	33.3	2	33.3
DRUS	85	12	85.9	16	81.2
DRXR	344	101	70.6	112	67.4
DSEC	96	33	65.6	32	66.7
DSEE	9	4	55.6	5	44.4
DSEM	1	0	100.0	0	100.0
DVDM	28	16	42.9	13	53.6
DVEN	20	5	75.0	5	75.0
DVOR	92	63	31.5	61	33.7
GCHE	19	12	36.8	14	26.3
GFIG	102	46	54.9	40	60.8
GFLO	20	13	35.0	13	35.0
GGEL	30	15	50.0	13	56.7
GGEN	21	16	23.8	11	47.6
GHDR	54	19	64.8	25	53.7
GMAT	5	5	0.00	5	0.00
GNCP	37	22	40.5	24	35.1
GPLI	22	10	54.5	3	86.4
GSCR	20	10	50.0	10	50.0
GSYS	16	7	56.3	9	43.8
GTAB	29	15	48.3	19	34.5
overall	2582	866	66.5	888	65.6

from in total 21 bounding boxes, the final dimension of feature space is 147. We call this approach BoBB-based approach. As compound image classification is a one-to-one classification problem, only the SVM classifier was applied on BoBB feature space. A grid searching on cost and value was used for parameter selection. Figure 3.1 shows the classification performance obtained by various BoBB strategies. Even color information was discarded, applying BoBB gray-level image (the value channel of image) can already obtain around 85% accuracy. Adding color features (approach 2), removing background (approach 3), enhancing connectivity (approach 4) can all improve the performance. Adding color or tensor features makes classifier more robust. Removing background eliminated false alarms, and archived the best improvement. Combining all these features, the best BoBB-based approach achieved 89.0% classification accuracy on training data and 85.1% on testing data using SVM classifier.

In non-compound classification, SURFContext + BoF approach was used. Both SVM classifiers and classical kNN classifiers were tested on BoF feature space. However, experiments on training data has shown that SVM classifiers significantly outperformed kNN classifiers (69.1% vs 48.1%). Therefore, only two runs using SVM classifiers were submitted to the modality classification task. Accuracy by class on training data shows most diagnosis image classes obtained over 80% accuracy, but on test data, the scores are around 70%, which shows the performance of this approach is not stable. Classes with large content diversity constantly obtained poor performance, Furthermore improvement is required on these classes.

3.2 Image-based Retrieval

For the image-based retrieval task, Mean average precision (MAP), binary preference (Bpref), and early precision (P10, P30) are shown as measures. Only one baseline (sari_SURFContext_HL_baseline) using was submitted and the corresponding MAP is 0.0086. Further tests on fusion strategies failed to improve the performance, because the runs to be fused themselves contain too few relevant results. The textual runs still largely outperforms the visual runs for both best score and average score. Results of our run and best runs of various nature are shown in Table 3.

Table 3. Results of the runs for the image-based topics.

Run	run_type	MAP	Bpref	P10	P30
best textual run (XRCE)	Textual	0.3196	0.2982	0.3886	0.2686
best visual run	Visual	0.0185	0.0361	0.0629	0.0581
sari_SURFContext_HL_baseline	Visual	0.0086	0.0181	0.0429	0.0352
best mixed run	Mixed	0.3196	0.2983	0.3886	0.2686

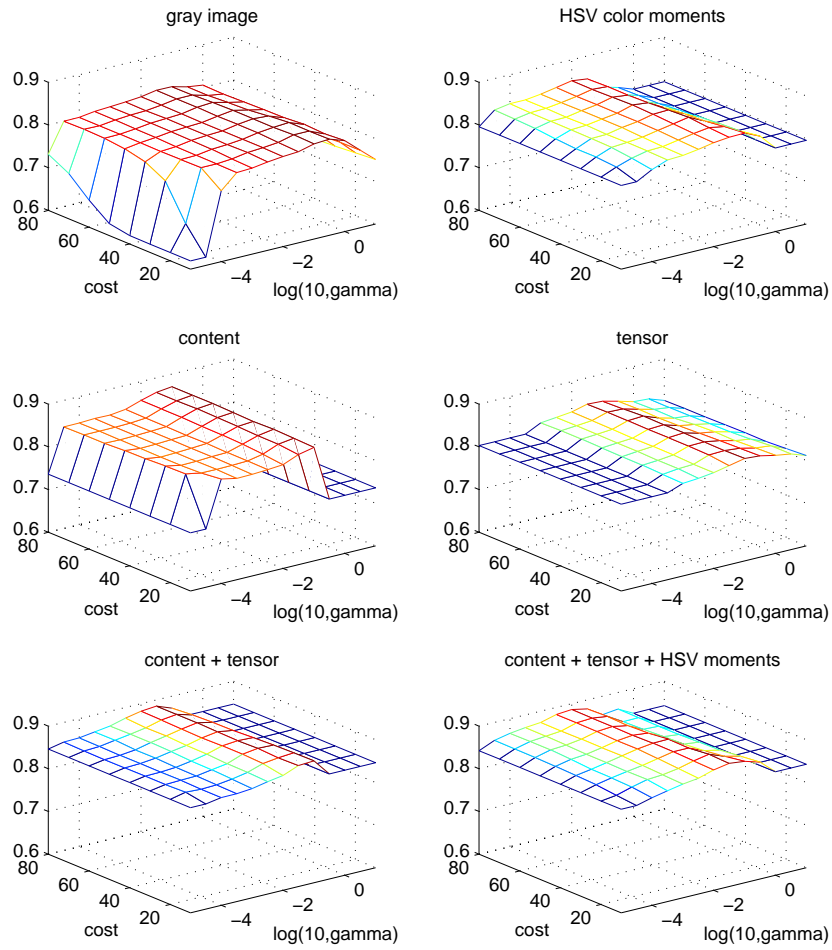


Fig. 1. The performance obtained with grid searching approach on training data. From left up to right bottom: 1) convert image to gray-level image and apply BoBB approach; 2) convert image to HSV space and apply BoBB approach; 3) using Laplace filter to remove background area, and then apply BoBB approach only on content region; 4) using structure tensor to enhance connectivity, and then apply BoBB approach; 5) combining 3) and 4); 6) combining 2), 3), and 4).

Apply fast filtering techniques to reduce the dimension and diversity of data was tried out. As most query images are diagnosis images, the goal is to reject those images which are completely not related to the query image in processing. Online query requires a quick indexing and search. Line/dot filter, Laplace filter, monochrome filter can process one image within 0.1 second. We used them to reject unwanted images from the huge database by the following order shown in Table 4. In our experiments, we define the region where standard derivation is

Table 4. Number of images rejected by fast filtering approach.

filtering	rejected	left
all line rejection	117'228	189'310
figure rejection	62'237	127'073
all text rejection		34'288 92'785

smaller than a threshold t . When all images were normalized to 0–255, t was set to be 5. Using line filter, the area of line-like structure can be obtained. If the area of whole image is equal to the sum of the area of line-like structure and the area of background, it is labeled as "all line". All line rejection aims to reject images containing only curves and lines. The risk of this filter is that it rejects also diagnostic printed signals, such as electrocardiography, etc. However, this is in general a safe solution, as query images are mainly diagnosis image.

Figure rejection aims to reject the man-made images, such as flowchart, system overviews, screenshots, etc. The Laplace filter was applied and where laplace equals 0 were located as "even" regions. If the area of whole image is equal to the sum of the area of line-like structure, the area of even regions and the area of background, it is labeled as "figures".

The third round, images containing only text, such as program listing, DNA sequences, tables and formulas, are rejected. This step is called "all text rejection". A orientation histogram was calculated on the line-like structures. Those images containing high repeatability in orientation histogram were rejected.

Finally, only 92'785 images are left for retrieval, 66% of data are quickly rejected. A comparison is made between the 92'785 images and the qrel file. There are 18'961 non identical image ids in qrel file, 17'415 of 18'961 (about 92%) are found in 92'785 images. All image labeled as "relevant" are not lost. In other words, fast filtering techniques reduced largely the diversity of content, but has little negative impact on image retrieval performance. This approach can be helpful to reduce the quantity of data for all the groups.

3.3 Conclusions

This paper describes the techniques used and results obtained by the MIILab group in ImageCLEFmed 2013. Two tasks are targeted: modality classification

and medical image retrieval. The techniques used by MIILab are purely visual-based. Classical BoF approach based on SURFContext was used as baselines, and several variants were tried to improve the performance. Bags of bounding boxes-based approach was proved to be useful for compound image classification. Fast filtering techniques were applied to reduce the scale of data, and was proved to be a safe strategy. For both modality classification and image retrieval tasks, MIILab was ranked in the middle among all the groups. This is the first year for MIILab to participate ImageCLEFmed, there is still room for performance improvement.

References

1. Clough, P., Müller, H., Sanderson, M.: The CLEF cross-language image retrieval track (ImageCLEF) 2004. In Peters, C., Clough, P., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B., eds.: *Multilingual Information Access for Text, Speech and Images: Result of the fifth CLEF evaluation campaign*. Volume 3491 of *Lecture Notes in Computer Science (LNCS)*, Bath, UK, Springer (2005) 597–613
2. Garcia Seco de Herrera, A., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., Müller, H.: Overview of the imageclef 2013 medical tasks. In: *Working notes of CLEF 2013*. (September 2013)
3. Caputo, B., Müller, H., Thomee, B., Villegas, M., Paredes, R., Zellhofer, D., Goeau, H., Joly, A., P., B., Martinez Gomez, J., Garcia Varea, I., Cazorla, M., eds.: *ImageCLEF 2013: the vision, the data and the open challenges*. In Caputo, B., Müller, H., Thomee, B., Villegas, M., Paredes, R., Zellhofer, D., Goeau, H., Joly, A., P., B., Martinez Gomez, J., Garcia Varea, I., Cazorla, M., eds.: *Proceedings of CLEF 2013. Lecture Notes in Computer Science (LNCS)*, Valencia, Spain, Springer (2013)
4. Li, Q., Sone, S., Doi, K.: Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional ct scans. *Medical Physics* **30**(8) (2003) 2040–2051
5. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medicine—clinical benefits and future directions. *International Journal of Medical Informatics* **73**(1) (2004) 1–23
6. Smith, J.R., Chang, S.F.: VisualSEEK: a fully automated content-based image query system. In: *The Fourth ACM International Multimedia Conference and Exhibition*, Boston, MA, USA (November 1996)
7. Squire, D.M., Müller, W., Müller, H., Pun, T.: Content-based query of image databases: inspirations from text retrieval. *Pattern Recognition Letters (Selected Papers from The 11th Scandinavian Conference on Image Analysis SCIA '99)* **21**(13–14) (2000) 1193–1198 B.K. Ersboll, P. Johansen, Eds.
8. Zhou, X.: *Grid-based Medical Image Retrieval Using Local Features*. PhD thesis, University of Geneva (2011)
9. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). *Computer Vision and Image Understanding* **110**(3) (2008) 346–359
10. Belongie, S., Greenspan, H., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence* **24**(4) (2002)
11. Zhou, X., Depeursinge, A., Müller, H.: Information fusion for combining visual and textual image retrieval. In: *20th IEEE International Conference on Pattern Recognition (ICPR)*. (August 2010) 1590–1593

12. Zhang, S., Yang, M., Cour, T., Yu, K., Metaxas, D.: Query specific fusion for image retrieval. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C., eds.: European Conference on Computer Vision. Lecture Notes in Computer Science. Springer Berlin Heidelberg (2012) 660–673