

Two-stage SVM for Medical Image Annotation

Bo Qiu, Changsheng Xu, Qi Tian
Institute for Infocomm Research (I²R)
21 Heng Mui Keng Terrace,
119613, Singapore
 {qiubo,xucs,tian}@i2r.a-star.edu.sg

ABSTRACT

In this paper, we proposed a two-stage medical image annotation method in order to achieve higher classification rate. Coarse and fine classifications are performed at the two stages. At the first stage, low resolution pixel maps (32x32) are used to represent the medical images while Support Vector Machines (SVM) are used to classify the resolution reduced pixel maps. Our experiment results showed that with the first stage classification 78.2% of classification accuracy has been achieved for the development dataset (DEV). At the second stage, selected images for which SVM classifiers are too close (judged by a predefined threshold value) are further refined in order to improve classification accuracy. In our experiments on DEV we have found there were about 200 images to be re-classified. Moreover, to eliminate the influence of the great volume unbalance among classes, a new training dataset is formed out of the old one. In this new balanced training dataset, each class has 30 samples at the most and the smallest class has 10 samples. We have further designed the following three classification steps: 1) 20x50 low resolution pixel maps feed into SVMs; 2) SIFT features feed into Euclidean distance classifiers; 3) 16x16 low resolution pixel maps feed into PCA classifiers. In the sequential steps, those classified classes which have little error are recorded (proven by experiments on DEV) and their results are taken as the last classification results instead of the initial results coming from the first SVM process. Finally the results from the SVMs at the first stage, and the refined classification results at the second stage are combined to form the final classification result. Our experimental results showed with the two stage method significant improvement of classification accuracy has been achieved.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing — *Abstracting methods, Dictionaries, Indexing methods*

General Terms

Algorithms, Performance, Experimentation

Keywords

Classification, Medical images, SVM, Two-stage

1. INTRODUCTION

Medical image annotation has great potential commercial value. However by now, the relevant techniques still lingers in a basic research stage. To boost the development of it, after the establishment of ImageCLEF, a sub-branch of CLEF (Cross Language Evaluation Forum) workshop (<http://www.clef-campaign.org/>), Human Language Technology and Pattern Recognition Group of the RWTH Aachen University helped to design a task of medical image annotation.

It's the second year that ImageCLEF publishes the medical image annotation task. Still at a starting stage, the annotation task is simplified to a classification problem: purely using visual features to classify some unlabeled images into some defined classes which are given by doctors. The total process is demanded to be automatic.

In ImageCLEF2006 there are 116 classes to be classified (Figure 1), which is totally different from 57 classes in ImageCLEF2005. Another difference is that there is an extra DEV provided which includes 1000 labeled images. The common parts of 2005 and 2006 are: 9000 labeled training images, 1000 testing images. And the evaluation method is simple:

to minimize the number of misclassified images in the 1000 testing data.

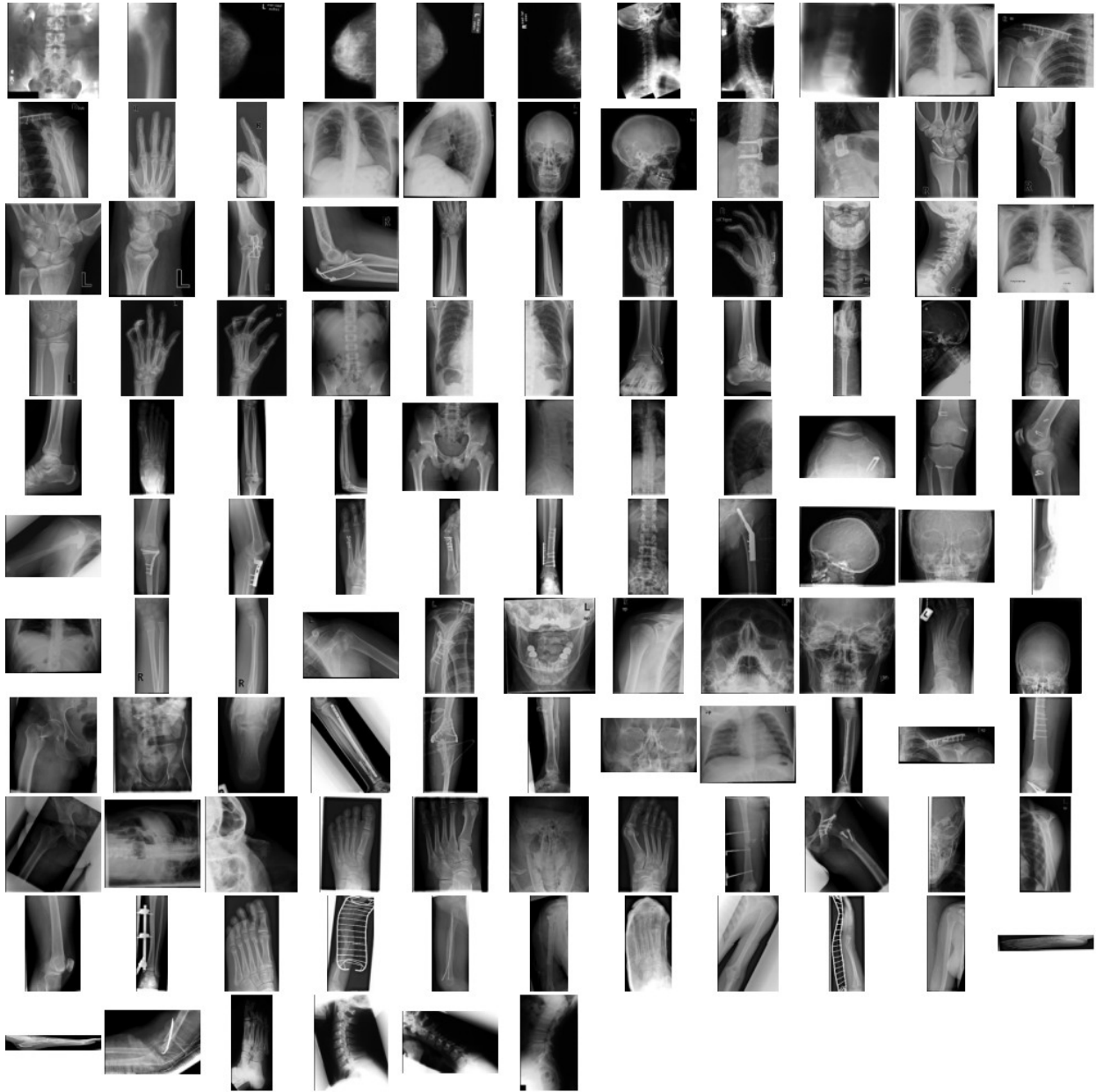


Figure 1. 116 classes of the task

In general each solution for the classification problem will include how to choose features and how to construct the corresponding classifiers. By now, there are many features and classifiers having been explored. Wherever, how to find the most efficient combination of the features and the classifier are the key factors to the solution. Features are described in Section 2; classification method is described in Section 3; and our result in joining the ImageCLEF2006 is shown in Section 4; at last the conclusion and future work is shown in Section 5.

2. IMAGE FEATURES

According to [1][2][3], the most frequently used image features are categorized into color, texture, and shape. In our past work [4], we tested Low Resolution Pixel-Map (LRPM) which can be regarded as a kind of color feature; besides of LRPM, we also explored contrast, anisotropy, polarity, which are texture features, and Blob, a mid-level feature compared with those frequently used low-level features. All the features are shown in Figure 2.

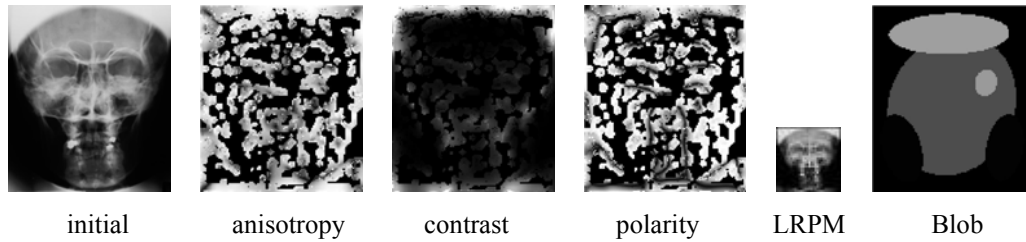


Figure 2. Features used in last year

In this year, we tested some new features (as shown in Figure 3) like salient map, salient point, SIFT (Scale Invariant Feature Transform), and stripe.

Salient map is an explicit two-dimensional map that encodes the saliency or conspicuity of objects in the visual environment [5][6]. “The purpose of the saliency map is to represent the conspicuity—or “saliency”—at every location in the visual field by a scalar quantity and to guide the selection of attended locations, based on the spatial distribution of saliency.” The last map is 16x16 and forms a 256 feature vector with the gray values.

Instead of global features like LRPM and salient map, in [7] and [8], the term ‘salient point’ is introduced, which is different from ‘interest point’ calculated by ‘corner detectors’ often although both of them are based on local computation. Salient points are related to any visual interesting part of the image whether it is smoothed or corner-like, while they are extracted under multi-resolution representation from wavelets. According to [9], choose the 50 top salient points and generate the image patches around the points. Our patches are 13x13 and gray values of those patches are formed into the feature vectors.

SIFT was introduced by David G. Lowe [10], and there are 4 major stages of computation used to generate SIFT features: Scale-space extrema detection, keypoint localization, orientation assignment, keypoint descriptor. SIFT transforms image data into scale-invariant coordinates relative to local features. For each keypoint, 4x4 descriptors are computed from a 16x16 sample array, and for each descriptor, there are 8 orientation bins. Thus a 128 element feature vector is formed.

Stripe was developed in our recent work [11]. By dividing an image into different kinds of grids, we can extract features based on the cells of the grids. To make it easily handled, histogram of each cell is calculated and we call the cells ‘stripes’. For one image, all its stripes’ histogram queues in order and forms a feature vector.

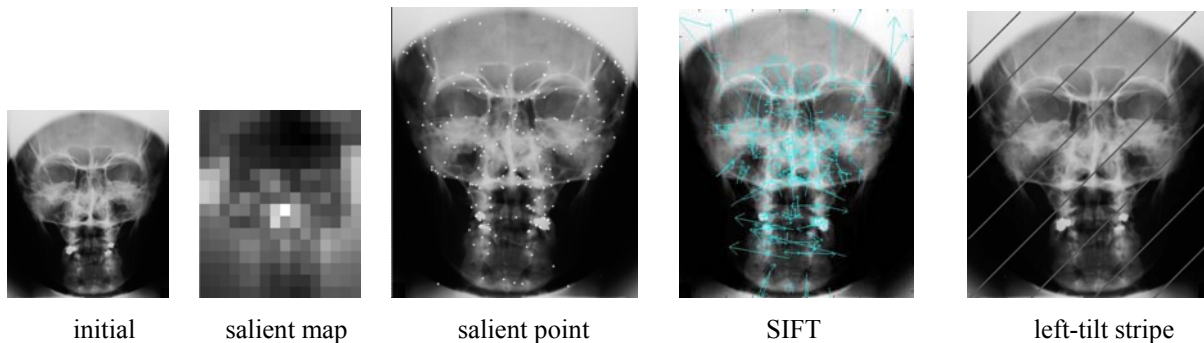


Figure 3. Features tested in this year

Facing so many features, we can not judge which one or which combination is the best for the solution of the task. They only can be chosen by the experiments. Besides of how to choose image features, another important factor influencing the classification results is how to design the classifiers.

3. CLASSIFICATION METHODOLOGY

According to [12], there have been many classifiers developed. And they can be basically divided into two categories: rule-based classifiers and computational intelligence based classifiers. The former one needs rules defined by designers, while the latter is just a framework made by designers and integrated with a learning algorithm, which includes supervised and

unsupervised methods. Unsupervised methods are also called clustering. Supervised methods include Boosting, Decision trees, Neural Networks, Bayesian Networks, SVM, Hidden Markov Field, etc.¹

In our system, SVM is chosen as the classifier. Different from other methods, we use a two stage SVM in the system. In the following, we firstly introduce a basic system using SVM. Then the two stage SVM is described.

3.1 A Basic Classification System Using SVM

As shown in Figure 4, SVM is a kind of supervised method. So the learning process is needed, in which the main target is to tune the parameters. After the parameters are fixed by the cross validation part, the unlabeled images can be input and processed by the classification system.

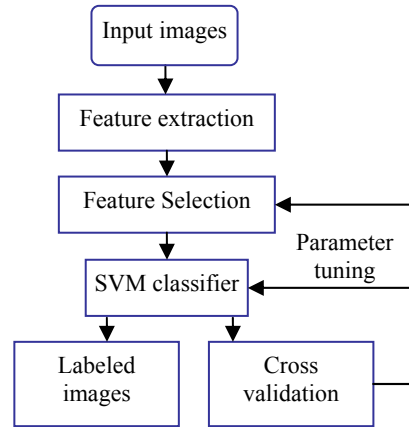


Figure 4. A basic classification system

SVM classifier can calculate similarity values between each input image and each class. Then for each input image, there will be a distance vector whose elements represent the distance values between the image and each class boundary. The smaller it is, the more difficult to be classified. So the largest value puts the image into the corresponding class.

Cross validation part uses some training date to tune the parameters. When using radial basis function (RBF), SVM parameters include the standard variance (σ), and the trade-off between training error and the margin (C) [4].

3.2 Two-stage SVM

The basic classification system using SVM can not reach a high accuracy if the image features are not well extracted. To improve the accuracy of the basic system, we put forward the two-stage SVM. The main idea is to find some rules to post-process the basic SVM results. So the two-stage SVM can be described as the following steps:

- 1) Execute the basic SVM classification process;
- 2) Find some badly classified classes and refine the results coming from step 1).

In the 2nd process, how to find the badly classified classes is based on the cross validation part. Owing to the seriously unbalanced problem (the numbers of the 116 classes are not equal but having great difference, the biggest class has more than 2500 samples while the least one has less than 10 samples), most of the wrongly classified samples focus on several classes. Those classes are proved by cross validation part that they include more than 95% wrongly classified samples. Targeting on refining the results of those classes, the 2nd step is presented in the following paragraph.

For one image in those chosen ‘bad’ classes, note that a corresponding distance vector from the basic system is $\mathbf{D} = (d_1, d_2, d_3, \dots, d_n)$. Inside the n elements, the maximal 3 values are $d_{k_1}, d_{k_2}, d_{k_3}$. Then if

$$(d_{k_1} - d_{k_2}) \times 10000 \leq th_1, \tag{1}$$

the image should be considered to be reclassified. th_1 is a threshold which is decided by the cross validation part.

¹ http://en.wikipedia.org/wiki/Statistical_classification

However, for all the images chosen by this way, there are around 1/3 of them are correctly classified, which has been proved by the cross validation experiment. This means the 1/3 needn't to be reclassified. How to identify those images? We find that if the image distance vector satisfies the following condition

$$(d_{k_2} - d_{k_3}) \times 10000 > th_2, \quad (2)$$

the image will be regarded as a 'good' image and needn't be reclassified. th_2 is also a threshold decided by the cross validation part.

In a short summary, to find the 'bad' images in the 2nd step of our two-stage SVM, firstly cross validation helps find the 'bad' classes; secondly the two formulas help judge whether the images should be reclassified or not.

While keeping the classified results of all the 'good' images, we are going to reclassify the chosen 'bad' images. This refining process includes the following steps:

- 1) Generating a new training dataset out of the old one. This is to eliminate the influence of the great volume unbalance among classes. In this new balanced training dataset, each class has 30 samples at the most and the smallest class has 10 samples.
- 2) Feeding 20x50 LRPM into SVM classifiers;
- 3) Feeding SIFT features into Euclidean distance classifiers;
- 4) Feeding 16x16 LRPM into PCA classifiers [4];
- 5) Choosing the best classifier and features for each class using cross validation;
- 6) Classifying the 'bad' images using the classifiers and features fixed by step 5).

At last, the results of the reclassified 'bad' images are combined together with those of the 'good' images coming from the first SVM process.

4. OUR RESULT FOR IMAGECLEF2006 MEDICAL ANNOTATION

In the case of fulfilling the task of ImageCLEF2006, we use DEV to make the cross validation to train the classifiers and choose image features. Then the 1000 true images are tested based on the trained classifiers.

4.1 Cross validation

As shown in Figure 5, when using 32x32 LRPM and SVM classifiers, our system reached an average accuracy of 78.2%. For all of the other features as mentioned in Section 2, we can not get a higher accuracy. So only this result is presented.

In the figure, the first part is 'correctness-class', which means the vertical axis presents the correctness/accuracy, and the horizontal axis presents the 116 classes. From this part we can see which classes have reached a satisfied accuracy. The second part shows the number of wrongly classified samples for each class. In this part we can find which class is the most difficult class. The difficult classes absorbed too many samples which don't belong to them. The third part shows the numbers of the classified images for each class and the fourth part shows the true numbers of images in each class.

In the whole process of cross validation, the SVM parameters are set and the two thresholds in formula (1) and (2) are chosen:

$$th_1 = 1.37,$$

$$th_2 = 3.$$

4.2 True test

In the first stage of SVM, 32x32 LRPM is used, the same as used in cross validation. The classification result is shown in Figure 6. The average accuracy accounting for all the 116 classes is 70.1%, which is calculated based on the published ground truth from the ImageCLEF2006med.

In the second stage of SVM, we used the features and classifiers as mentioned in Section 3.2. Combined with the results from the first stage, the final result is shown in Figure 7. The average accuracy over the 116 classes is 72%.

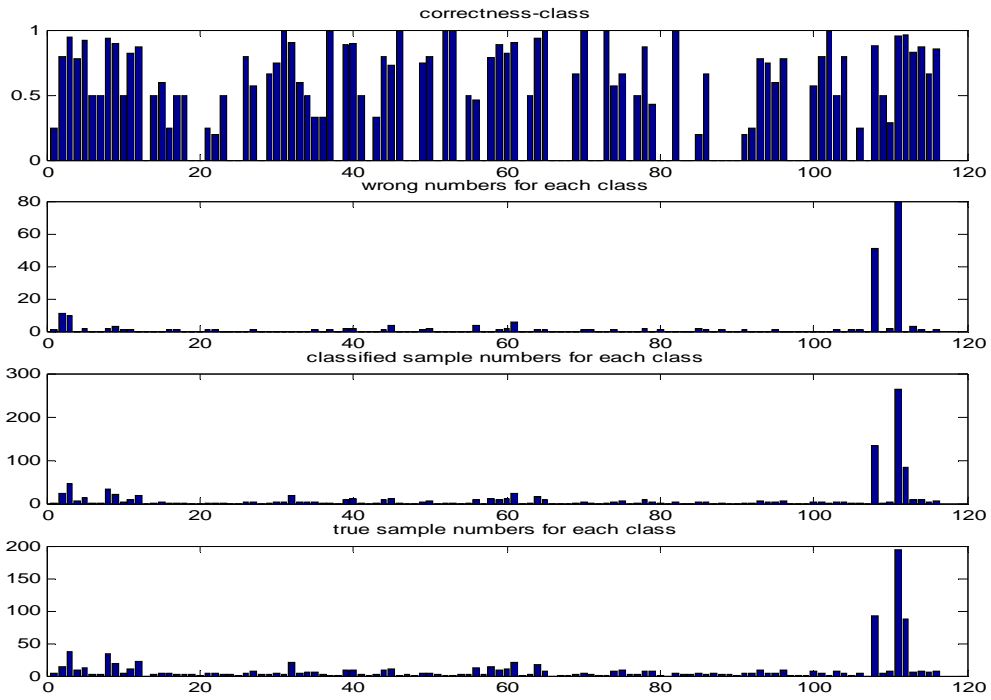


Figure 5. Result of cross validation using DEV (accuracy = 78.2%)

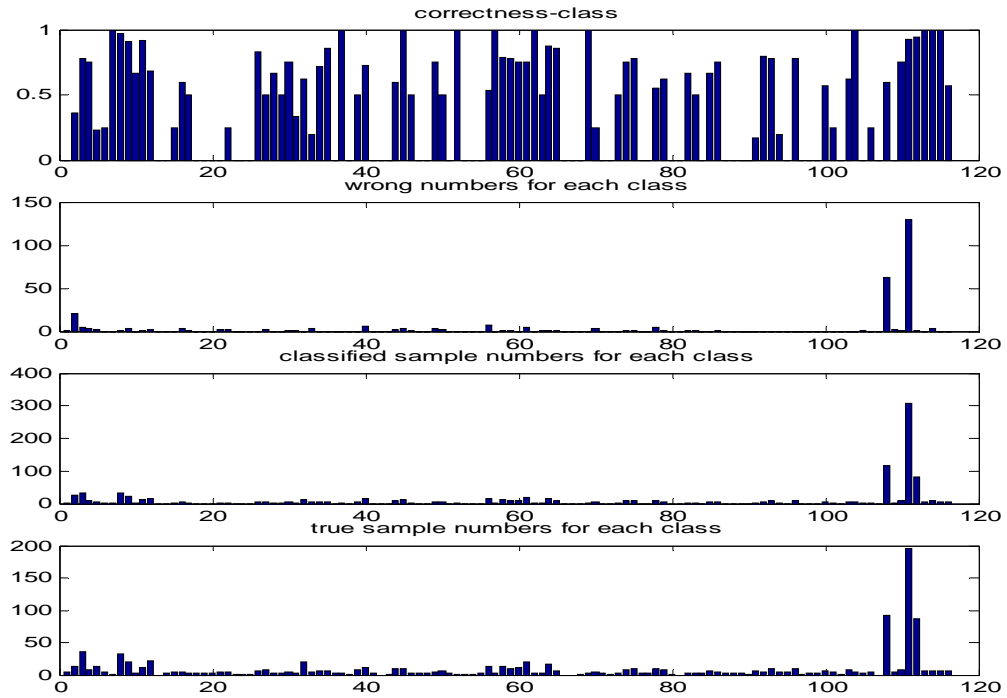


Figure 6. Initial result using first stage SVM (accuracy = 70.1%)

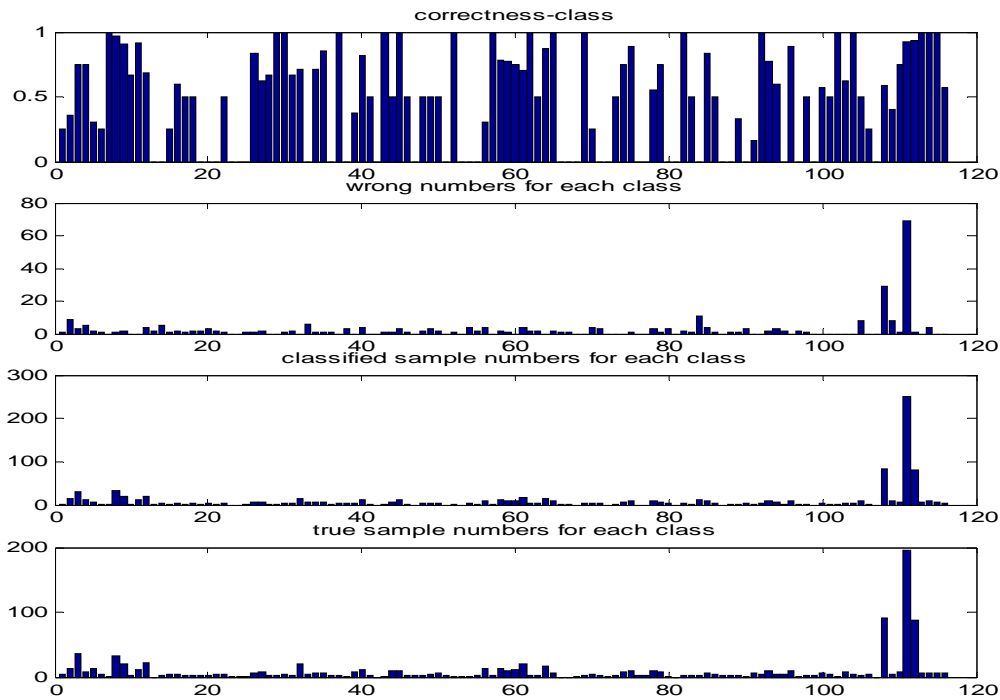


Figure 7. Result after refining with two-stage SVM (accuracy = 72%)

5. CONCLUSION AND FUTURE WORK

Though we got a common result in DEV, the true results with the 72% accuracy is too far away from expected. The main reason for this may be owing to the bad choice of features. Global features are not good enough to be applied to solve the complex classification problem of medical images. More precise local features have to be considered.

In the future work, we will focus on more efficient features and relevant classifiers. And more work on analyzing the 116 classes should be done because many classes can not be identified even with our eyes. Where is the key part/feature to distinguish them? It may be a long-term question.

6. REFERENCES

- [1] Remco C. Veltkamp, Mirela Tanase, Content-Based Image Retrieval Systems: A Survey, Technical Report UU-CS-2000-34, Oct. 2000, <http://give-lab.cs.uu.nl/cbirsurvey/>.
- [2] T. Lehmann et al, Automatic Categorization of Medical Images for Content-based Retrieval and Data Mining, *Computerized Medical Imaging and Graphics*, Vol. 29, pp. 143-155, 2005.
- [3] Björn Johansson, A Survey on: Contents Based Search in Image Databases, LiTH-ISY-R-2215, Technical Reports from the Computer Vision Laboratory, Dept. of Electrical Engineering, Linköping University, Sweden, Feb., 2000.
- [4] Bo Qiu, Changsheng Xu, Qi Tian, An Automatic Classification System Applied in Medical Images, IEEE Intl Conf Multimedia & Expo (ICME) 2006, pp.1045-1048, Toronto, Canada., Jul. 9-12, 2006.
- [5] L. Itti, C. Koch, E. Niebur, A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, pp. 1254-1259, Nov 1998.
- [6] L. Itti, C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research*, Vol. 40, No. 10-12, pp. 1489-1506, May 2000.
- [7] N. Sebe, Q. Tian, E. Loupias, M.S. Lew, T.S. Huang, Evaluation of Salient Point Techniques, *Image and Vision Computing*, Vol. 21, No. 13-14, pp. 1087-1095, December, 2003.
- [8] E. Loupias, N. Sebe, S. Bres, J-M. Jolion, Wavelet-Based Salient Points for Image Retrieval, *International Conference on Image Processing (ICIP'00)*, Vancouver, Canada, 2000.
- [9] T.Deselaers, D.Keysers, H.Ney, Improving a Discriminative Approach to Object Recognition using Image Patches,

Proc. DAGM 2005, LNCS 3663, pp. 326-333, Vienna, Austria, Springer.

- [10] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, No.2, pp. 91-110, 2004.
- [11] Bo Qiu, Daniel Racoceanu, Chang Sheng Xu, and Qi Tian, Stripe: Image Feature Based on a New Grid Method and Its Application in ImageCLEF, will appear in AIRS 2006, LNCS 4182 proceedings.
- [12] Romesh Ranawana, Vasile Palade, "Multi Classifier Systems – A Review and Roadmap for Developers", to appear in March 2006, Issue of the Journal of Hybrid Intelligent Systems, IOS Press Amsterdam.