

Automatic Annotation of X-ray Images: A Study on Attribute Selection

Devrim Unay¹, Octavian Soldea¹, Ahmet Ekin², Mujdat Cetin¹, and Aytul Ercil¹

¹ Computer Vision and Pattern Analysis Laboratory, Faculty of Engineering and Natural Sciences, Sabanci University, Turkey, unay@sabanciuniv.edu,

² Video Processing and Analysis Group, Philips Research, The Netherlands.

Abstract. Advances in the medical imaging technology has lead to an exponential growth in the number of digital images that need to be acquired, analyzed, classified, stored and retrieved in medical centers. As a result, medical image classification and retrieval has recently gained high interest in the scientific community. Despite several attempts, the proposed solutions are still far from being sufficiently accurate for real-life implementations.

In a previous work, performance of different feature types were investigated in a SVM-based learning framework for classification of X-Ray images into classes corresponding to body parts and local binary patterns were observed to outperform others. In this paper, we extend that work by exploring the effect of attribute selection on the classification performance. Our experiments show that principal component analysis based attribute selection manifests prediction values that are comparable to the baseline (all-features case) with considerably smaller subsets of original features, inducing lower processing times and reduced storage space.

1 Introduction

Storing, archiving and sharing patient information among medical centers has become a crucial task for the medical field. Companies as well as governments are now in anticipation of building Patient Centric IT systems such as that targeted by the Ratu e-health project of Northern Finland³ that focus on building a large national electronic patient records archive.

Digital medical images, such as standard radiographs (X-Ray) and computed tomography (CT) images, represent a huge part of the data that need to be stored in medical centers. Manual labeling of this data is not only time consuming, but also error-prone due to inter/intra-observer variations. In order to realize an accurate classification one needs to develop tools that allow high performance automatic image annotation, i.e. labeling of a given image with some text or code without any user interaction.

³ <http://pre20090115.stm.fi/pr1105954774022/passthru.pdf>

Several attempts in the field of medical images have been performed in the past. For example, the WebMRIS system [1] aims at retrieving cervical spinal X-Ray images, whereas the ASSERT system [2] focuses on retrieving CT images of the lungs. While these efforts consider retrieving a specific body part only, other initiatives have been taken in order to retrieve multiple body parts.

The ImageCLEF Medical Image Annotation task, run as part of the Cross-Language Evaluation Forum (CLEF) campaign, is a yearly held medical image annotation challenge for automatic classification of an X-Ray image archive containing more than 10,000 images randomly taken from the medical routine. The ImageCLEF Medical Annotation dataset contains images of different body parts of people from different ages, of different genders, under varying viewing angles and with or without pathologies. Depending on the year of the challenge, participants are asked to automatically annotate these images according to classification labels that vary from 58 to 196.

A potent classification system requires the image data to be translated into a more compact and more manageable representation containing only the relevant features. Several feature representations have been investigated in the past for such a classification task. Among others, image features, such as average value over the complete image or its sub-regions [3] and color histograms [4], have been investigated. Recently in [5], texture features such as local binary patterns (LBP) [6] have been shown to outperform other types of low-level image features in classification of X-Ray images. One drawback of the mentioned work is the large number of features extracted, which may be problematic for the classification step. Retaining only the relevant features by applying attribute selection on local binary patterns, may lead to comparable classification accuracies with smaller feature sets.

Motivated by the considerations above, in this paper we explore the effect of principal component analysis based feature selection on the performance of local binary patterns applied to the ImageCLEF-2009 Medical Annotation dataset.

The paper is organized as follows. Section 2 presents our feature extraction, feature selection and classification steps in detail. Section 3, introduces the image database and the experimental evaluation process performed. Next, in Section 4, corresponding results are presented. Finally, Section 5 outlines our conclusion.

2 Method

In this work we utilize the image data from the ImageCLEF-2009 Medical Annotation task for training and testing. 12677 fully classified and unbalanced X-Ray images are available to train a classification system, which needs to be evaluated using 2000 unlabeled images according to four different label sets including 58 to 196 distinct classes. Please note that, the data is unbalanced meaning some classes have significantly larger share among data than others.

2.1 Feature Extraction

We extract *spatially enhanced local binary patterns* as features from each image in the database. LBP [6] is a gray-scale invariant local texture descriptor with low computational complexity. The LBP operator labels image pixels by thresholding a neighborhood of each pixel with the center value and considering the results as a binary number. The neighborhood is formed by a symmetric neighbor set of P pixels on a circle of radius R . Formally, given a pixel at (x_c, y_c) , the resulting LBP code can be expressed in the decimal form as follows :

$$LBP_{P,R}(x_c, y_c) = \sum_{n=0}^{P-1} s(i_n - i_c) 2^n \quad (1)$$

where n runs over the P neighbors of the central pixel, i_c and i_n are the gray-level values of the central pixel and the neighbor pixel, and $s(x)$ is 1 if $x \geq 0$ and 0 otherwise.

After labeling an image with a LBP operator, a histogram of the labeled image $f_l(x, y)$ can be defined as

$$H_i = \sum_{x,y} I(f_l(x, y) = i), \quad i = 0, \dots, L - 1 \quad (2)$$

where L is the number of different labels produced by the LBP operator, and $I(A)$ is 1 if A is true and 0 otherwise.

The derived LBP histogram contains information about the distribution of local micro-patterns, such as edges, spots and flat areas, over the image. Following [6], not all LBP codes are informative, therefore we use the uniform version of LBP and reduce the number of informative codes from 256 to 59 (58 informative bins + one bin for noisy patterns). Following [7], we divide the images into 4x4 non-overlapping sub-regions and concatenate the LBP histograms extracted from each region into a single, spatially enhanced feature histogram (Figure 1). This step aims at obtaining a more local description of the image.

Finally, we obtain a total of 944 features per image, which is a large number for the classification step. Therefore, we apply principal component analysis based feature selection.

2.2 Feature Selection: Principal Component Analysis

Principal component analysis (PCA, or Karhunen-Loeve transform) [8] is a vector space transformation often used to reduce multidimensional datasets to lower dimensions for analysis.

Given data X consisting of N samples, in PCA we first perform data normalization by subtracting the mean vector m from the data. Then the covariance matrix Σ of the normalized data $(X - m)$ is computed.

$$m = \frac{1}{N} \sum_{i=1}^N X_i \quad (3)$$

$$\Sigma = (X - m)(X - m)^T \quad (4)$$

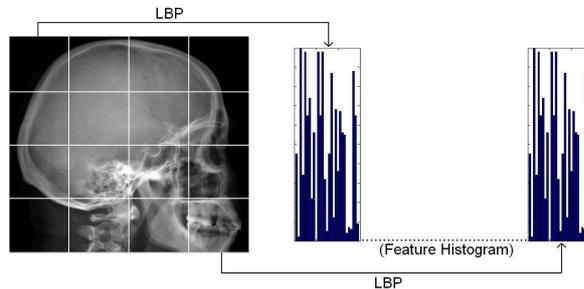


Fig. 1. The image is divided into 4x4 non-overlapping sub-regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced histogram.

Afterwards, the basis functions are obtained by solving the algebraic eigenvalue problem

$$\Lambda = \Phi^T \Sigma \Phi \quad (5)$$

where Φ is the eigenvector matrix of Σ , and Λ is the corresponding diagonal matrix of eigenvalues. Feature selection is then performed by keeping q ($q < N$) orthonormal eigenvectors corresponding to the first q largest eigenvalues of the covariance matrix. Here, q is empirically set such that total variance measured from these eigenvalues correspond to a user-defined percentage.

2.3 Classification: Support Vector Machines

SVM [9] is a popular machine learning algorithm that provide good results for general classification tasks in the computer vision and medical domains: e.g. nine of the ten best models in ImageCLEFmed 2006 competition were based on SVM [10]. In a nutshell, SVM maps data to a higher-dimensional space using kernel functions and performs linear discrimination in that space by simultaneously minimizing the classification error and maximizing the geometric margin between the classes.

Among all available kernel functions for data mapping in SVM, Gaussian radial basis function (RBF) is the most popular choice, and therefore it is used here.

$$RBF : K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0 \quad (6)$$

where γ is a parameter defined by the user. Besides γ , there exists an error cost C that controls the trade-off between allowing training errors and forcing rigid margins. An optimum C value creates a soft margin while permitting some misclassifications. In this work we used LibSVM library (version 2.89) [11] for SVM and empirically found its optimum parameters (γ and C) on the dataset.

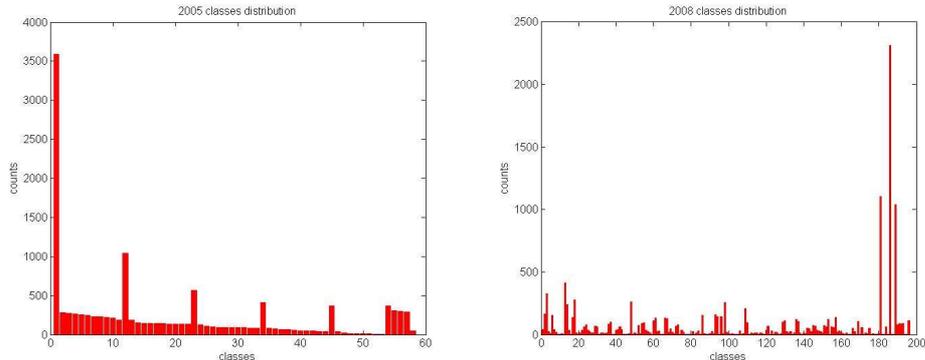


Fig. 2. Distribution of the data labeled in 2005 (left) and 2008.

3 Experimental Setup

3.1 Image Data

The training database released for the ImageCLEF-2009 Medical Annotation task includes 12677 fully classified (2D) radiographs that are categorized into 57 classes in 2005 and 196 classes in 2008. Their distribution with respect to these classes is displayed in Figure 2.

3.2 Evaluation

In order to avoid domination of attributes with greater numeric ranges over small ones, we linearly scale each feature to $[-1,+1]$ range before presenting them to the SVM.

We evaluate our SVM-based learning using 5-fold cross validation, where the database is partitioned into five subsets. Each subset is used once for testing while the rest are used for training, and the final result is assigned as the average of the five validations. Note that for each validation all classes were equally divided among the folds. We measure the overall classification performance using accuracy, which is the number of correct predictions divided by the total number of images.

4 Results

We compare our classification results of SVM with PCA-based feature selection (referred to as *SVMrbf+PCA* in the Figures 3-4) with two reference approaches: 1) *baseline (No PCA)* that refers to the SVM classification with all available features (PCA is not applied), and 2) *random guess* meaning the classifier puts all the data to the class with the highest frequency.

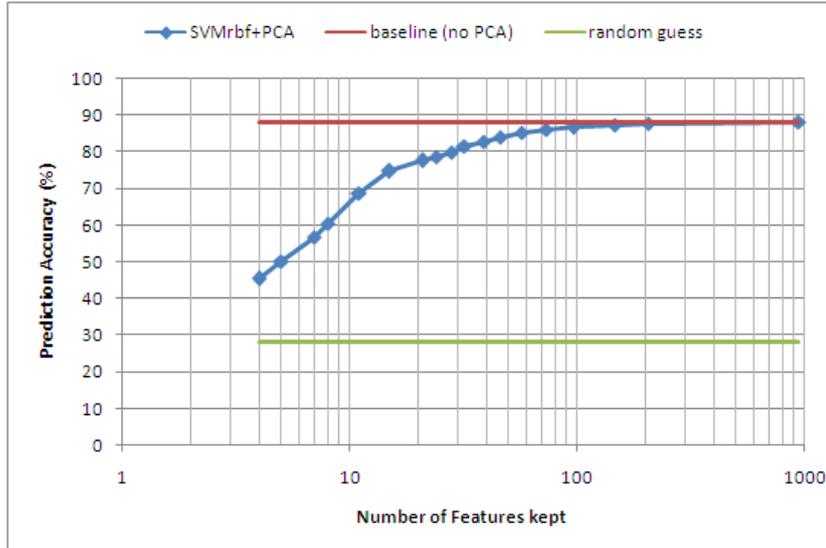


Fig. 3. Effect of PCA on classification accuracy with LBP features and 2005 labels used (57 classes).

Figure 3 shows the effect of PCA-based feature selection on the classification accuracy of SVM for the data with 2005 labels (57 classes). Notice that when all the LBP feature set is input to the SVM (baseline case), the overall accuracy is measured as 88%, while random guess is at the level of 28%. On the other hand, with attribute selection we reach accuracy levels (87,5%) comparable to the baseline case with only about 150-200 features out of possible 944. This leads to considerable reduction in prediction time as well as storage space. This observation shows that although the used LBP^{u2} operator inherently discards non-informative patterns from the feature set, we can further refine these attributes using PCA without degrading classification accuracy.

Figure 4 shows the effect of PCA-based feature selection on the classification accuracy of SVM for the data with 2008 labels (196 classes). For this case, baseline accuracy is measured as 83,4%, while random guess is at the level of 18%. Similar to the previous observations, performing feature selection with PCA results in accuracy values (83%) comparable to the baseline with approximately 150-200 features.

In terms of computational expense (on a PC with 2.13GHz processor and 2GB RAM), the baseline approach exhibits 25.5min processing time with 151MB storage space required for the cross-validation task. On the other hand, for the proposed PCA-based approach these values are measured as 4.4min and 23MB, respectively. In consequence, the proposed PCA-based approach provides an over 5-fold improvement in processing time and storage space requirements.

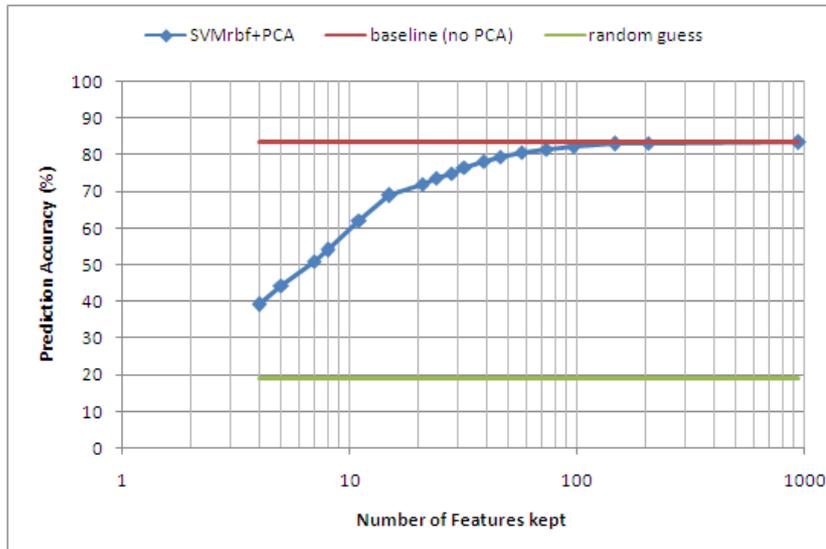


Fig. 4. Effect of PCA on classification accuracy with LBP features and 2008 labels used (196 classes).

5 Conclusion

In this paper we have introduced a classification work with the aim of automatically annotating X-Ray images. We have explored the effect of PCA-based feature selection on the efficacy of recently popular and highly discriminative local binary patterns within a SVM-based learning framework. Our experiments on the ImageCLEF-2009 Medical Annotation database revealed that applying attribute selection on local binary patterns provide comparable classification accuracies with considerably smaller number of features, leading to reduced processing time and storage space requirements.

References

1. Long, L.R., Pillemer, S.R., Lawrence, R.C., Goh, G.H., Neve, L., Thoma, G.R.: WebMIRS: web-based medical information retrieval system. In Sethi, I.K., Jain, R.C., eds.: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Volume 3312. (December 1997) 392–403
2. Shyu, C.R., Brodley, C.E., Kak, A.C., Kosaka, A., Aisen, A.M., Broderick, L.S.: Assert: a physician-in-the-loop content-based retrieval system for hrct image databases. *Comput. Vis. Image Underst.* **75**(1-2) (1999) 111–132
3. Rahman, M.M., Desai, B.C., Bhattacharya, P.: Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion. *Computerized Medical Imaging and Graphics* **32**(2) (2008) 95 – 108

4. Mueen, A., Sopian Baba, M., Zainuddin, R.: Multilevel feature extraction and x-ray image classification. *J. Applied Sciences* **7**(8) (2007) 1224–1229
5. Jacquet, V., Jeanne, V., Unay, D.: Automatic detection of body parts in x-ray images. In: *Mathematical Methods in Biomedical Image Analysis, 2009. MMBIA 2009. IEEE Computer Society Workshop on.* (2009)
6. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(7) (2002) 971–987
7. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: *Lecture Notes in Computer Science : Computer Vision - ECCV 2004.* (2004) 469–481
8. Jolliffe, I.T.: *Principal Component Analysis.* Second edn. Springer (October 2002)
9. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* **2**(2) (1998) 121–167
10. Müller, H., Deselaers, T., Deserno, T., Clough, P., Kim, E., Hersh, W.: Overview of the imageclefmed 2006 medical retrieval and medical annotation tasks. In: *Evaluation of Multilingual and Multi-modal Information Retrieval.* (2007) 595–608
11. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.