Lung Image Classification using Locality-Constrained Linear Coding

Meng Yi[1], Tatyana Nuzhnaya[1], Vasileios Megalooikonomou[1], Xinggang Wang[2], Longin Jan Latecki[1], Mark Kohn[4], Robert Steiner[3]

[1]Department of Computer and Information Science, Temple University, Philadelphia, PA, USA
[2]Department of Electronics and Information Engineering, Huazhong University of Science and Technology, China
[3]Department of Radiology, Temple University, School of Medicine, Philadelphia, PA, USA

Abstract. We propose a locality-constrained linear coding based approach for classifying lung images in a computed tomography (CT) image set. This method can be utilized to assess the tissue patterns in CT lung images, and thus assisting the diagnosis of Pulmonary Emphysema. Lung images in the set are divided into 4 categories corresponding to different lung diseases. The regions of interest (ROIs) are specified by domain experts. First, dense SIFT features are extracted out for each image; Then, we encode the features using a k-means generated codebook and locality-constraint linear coding (LLC), which is introduced in place of vector quantization (VQ) coding for it preserves both locally and sparsity. Next, two feature representations: spatial pyramid matching (SPM) and feature context (FC) are utilized to convey statistical as well as spatial information of lung images. The coded features are congregated and form a final descriptor vector of the image, which is fed into classifiers. Experiments demonstrate the advantage of our approach compared to baseline classification methods using VQ and histogram similarity. A 89.2% classification accuracy is achieved on a set of 200 lung images. It indicates that our approach can potentially aid in diagnosing lung diseases.

1 Introduction

Chronic obstructive pulmonary disease (COPD) is the occurrence of a pair of commonly co-existing diseases of the lungs: chronic bronchitis and emphysema. It is a growing health problem, due to an increase in smoking rates and demographic changes in many countries. COPD is now the third leading cause of death in the U.S, and it is predicted to be the fourth leading cause of death worldwide by 2030 [1]. Detecting and quantitatively analyzing emphysema are of great importance, since it is thought to be the main cause of disability in COPD.

High-resolution computed tomography(CT) imaging is an imaging technique for the assessment of patients with lung diseases. It is gaining more and more attention, because it provides more detailed information regarding lung parenchyma, compared to traditional diagnostic tools for COPD, for example, spirometry. A CT scan of the chest shows the distribution of emphysema throughout the lungs. It can be used to assess the pathological extent of emphysema [9]. In CT, emphysema can be divided into three subtypes: centrilobular emphysema (CLE), paraseptal emphysema (PSE) and interstitial lung diseases (ILD). The goal of this study is to classify images in an CT image dataset, which contains CT scans from the three subtypes above and a normal control group(NL).

One approach to characterize a lung image is using tree-like structures [7, 5, 10]. Another common approach is texture analysis [13, 8], in which local binary pattern-s(LBP) and histogram are usually utilized to capture texture information. Unlike these methods [12, 11], we combine texture features with feature representations which captures spatial distribution. Some [11] uses a coding strategy to reduce feature dimensionality. Vector quantization (VQ) is an often employed coding method. In this work, we introduce a new coding algorithm, which has theoretical advantage over VQ and shows promising performance in the experiments.

Our contribution is three folds: First, we introduce LLC coding to lung image classification, and demonstrate the performance boost compared to a baseline VQ coding method. Second, we proposed feature representations to combine traditional feature histogram with spatial layout, the benefit of this combination is proven in experiments. Third, a system for classifying lung disease types in CT images is proposed and a 89.2% classification accuracy is achieved.

The rest of the paper is organized as follows: Section 2 introduces our lung image classification method based on LLC coding. Experiments and evaluations are presented in Section 3. In Section 4 we conclude this paper.

2 Methods

An overview of our approach can be seen in Fig.1. First, Dense SIFT feature [4] is extracted for each image. This forms the initial high dimensional representation of each image. Second, to extract information and reduce dimensionality, we perform a coding step: All feature vectors are clustered into N piles using K-means algorithm. The center of each pile is selected as a "code". In the encoding phase, feature vectors can be re-formulated based on its relation to these codes. An exemplar coding algorithm is vector quantization (VQ) coding, in which each feature vector is defined by its n-earest code. Here we introduce locality-constrained linear coding(LLC) to convert the descriptors into codes. It represents each feature vector using a small number of nearby codes. More details about LLC can be found in Sec. 2.2. Next, we apply two kinds of feature representation strategies: spatial pyramid matching (SPM) and feature context (FC)[15] separately to concatenate feature vectors and generate the final representations of images. At last, the representations are fed into a classifier.

2.1 Feature Representation

Feature histograms are commonly used for characterizing CT lung images [12]. It is a compact representation which encodes the frequency of each feature in an image. However, it is incapable of recording spatial layout of the image, thus may fail to capture some differences between classes. In order to compensate for this shortcoming of feature histogram, we propose two kinds of feature representations: spatial pyramid

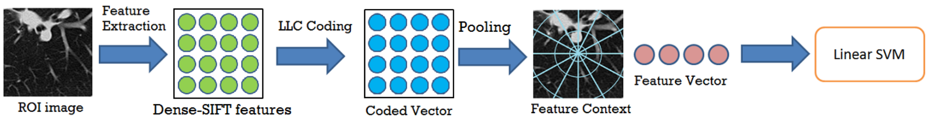

Fig. 1. Flowchart of our approach.The original image is first represented using Dense-SIFT features. Then the features are encoded using LLC coding. The coded features are pooled using Feature Context reorientation is shown in the figure, or spatial pyramid matching, then concatenated and fed in to a classifier.

matching (SPM) [11] and feature context (FC) [15]. Both can encode statistical information of the images, yet preserve some spatial distribution of features. Experimental results demonstrate the superior discriminative capacity of these feature representations compared to feature histogram.

Feature Context (FC) is motivated by a robust shape descriptor shape context (SC) [2]. The feature points are mapped into a log-polar coordinate system centered at some reference point. Each bin of the log-polar space is determined by angle and distance intervals. FC counts the feature points within each bin and represents the counts as a 2D histogram.

In a given image I, we have a set $Z = \{z_1, \dots, z_L\}$ of feature point locations. Each feature point $z \in Z$ is encoded as vector $C(z) = (w_1^z, \dots, w_K^z)$. Let p be a location of an reference point in I. Following SC, the area around p is divided into regions $Region_r^p$ in log polar coordinate system for $r = 1, ..., R$, e.g.,the feature context (FC) around point p is defined as a matrix

$$FC(p, r, i) = \mathcal{M}\{w_i^z \mid (z - p) \in \Delta_s(Region_r^p)\}, \tag{1}$$

where $i = 1, \dots, K$ indexes the codewords and $\mathcal{M}$ is a pooling function, which extracts the most relevant codewords present in the region $Region_r^p$. The function $\mathcal{M}$ can be max, sum, $mean$ or some other functions. We selected max pooling function as $\mathcal{M}$ in our experimental results, since the max pooling method is more robust to local transformation than mean statistics in histogram [16]. $\Delta_s(Region_r^p)$ denotes a neighborhood of region $Region_r^p$ of radius s. It allows us to compensate for spatial uncertainty of local descriptor. The local descriptors near the boundaries of regions may belong to multiple regions. By using $\Delta_s(Region_r^p)$ in Eq. (1), they are assigned to those regions, which increases the robustness of our Feature Context descriptor. In image I, we usually have a set of reference points as $P = \{p_1, p_2, ..., p_L\}$. Therefore, FC of an image I is a tensor of dimension $L \times R \times K$ given by

$$FC(I) = (FC(p, r, i))_{p,r,i} \in \mathbb{R}^{L \times R \times K} \tag{2}$$

Spatial Pyramid Matching SPM is first proposed in [3] for recognizing scene categories. It is a efficient and effective extension of bag-of-features (BoF) method like histograms, and has been proven to perform well in image classification tasks. The SPM method partitions image into increasingly finer spatial sub-regions, and computes histograms of local features for each sub-region. Typically, $2l \times 2l$ sub-regions, $l = 0, 1, 2$ are used. For example, an image is first viewed as a whole, then, it is segmented into 4 parts, and then every part is further segmented into 4 sub-parts. Fig.2 illustrates an example of SPM. A input image is shown on the top. Fig.2(a) is the feature histogram at the first layer. Fig.2(b) shows 4 histograms for the 4 subregions at the second layer. Fig.2(c) contains the 16 histograms for each of the 16 subregions at the third layer.

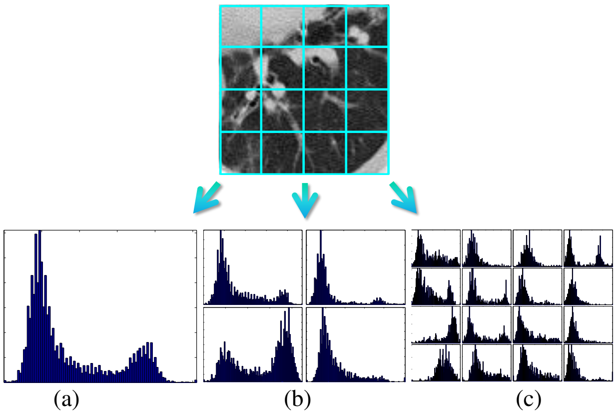

Fig. 2. Illustration of spatial pyramid matching. The image at the top is the input image. (a) (b) (c) are histograms at the three layers.

SPM approaches typically has two steps: First, local features (SIFT [6], LBP [12], etc.) are extracted from the image, forming the descriptor layer. Then, these features are quantized using a codebook and generate the code layer. In the SPM layer, multiple codes from inside each sub-region are pooled together by averaging and normalizing into a histogram. Finally, all histograms for sub-regions are concatenated to form the descriptor of the whole image.

2.2 Coding Strategy

Locality-constrained linear coding(LLC) [14] is an adaptation of sparse coding(SC) with locality constraints. It has several advantages over sparse coding and vector quantization(VQ).
Traditionally, vector quantization(VQ) is used to generate code from raw descriptors. It solves the following constrained least squrare fitting problem:

$$\begin{aligned} \arg\min_{c} \sum_{i=1}^{N} \|x_i - Bc_i\|^2 \\ s.t. \|c_i\|_{l^0} = 1, \|c_i\|_{l^1} = 1, c_i \geqslant 0, \forall i \end{aligned} \tag{3}$$

Where $X = [x_1, x_2, , x_n]$ is the descriptor of an image. B is a codebook. $C = [c_1, c_2, , c_n]$ is the set of code for image X. The optimization goal of VQ is to find a quantized code C with a single non-zero element which is approximately equal to X.

However, VQ method generates quantization loss, and is poor in scalability. In order to improve scalability and reduce quantization loss, ScSPM [16] method is proposed to introduce sparse coding to SPM procedure, obtaining nonlinear feature representation that work better with linear classifiers. Here the coding problem becomes a standard sparse coding problem:

$$\arg\min_{c} \sum_{i=1}^{N} \|x_i - Bc_i\|^2 + \lambda \|c_i\|_{l^1} \tag{4}$$

The l^0 restraint in Eq. (4) is relaxed by using a sparsity regularization term.

Instead of using sparsity constraint, in Locality-constrained Linear Coding(LLC), a locality constraint is incorporated into the optimization goal as follow:

$$\begin{aligned} \arg\min_{c} \sum_{i=1}^{N} \|x_i - Bc_i\|^2 + \lambda \|d_i \odot c_i\|^2 \\ s.t. 1^\top i = 1, \forall i \end{aligned} \tag{5}$$

Where x is the descriptor, B is codebook, and c is the code for this image. The second term in Eq.(5) denotes the element-wise multiplication. d_i is the locality adaptor that gives different freedom for each basis vector proportional to its similarity to the input descriptor x_i. An LLC procedure has three steps: First, for each input descriptor x_i, its K-Nearest Neighbors can be denoted as B_i. Then, x_i can be approximately reconstructed using the set of B_i. At last, representing the input descriptor x_i using the corresponding parameter c_i for each code in the codebook B. In this way, we use a vector to represent the input image, which is as large as the codebook, no matter what is the size of the extracted feature descriptors. This uniform description can then be fed into a SVM classifier and finally give out the image class predication.

The difference between VQ, SC and LLC coding is represented in Fig. 3. In VQ coding, each input is coded using only one most similar element from the codebook. This leads to large quantization error. While in SC and LLC, each input is represented by multiple elements from the codebook, which can better representing the inputs. Furthermore, by applying locality constraint, LLC captures the correlations between similar descriptors.

3 Experiments

3.1 Data Setting

The data come from a CT lung image dataset collected at Temple University Hospital. The Regions of Interest (ROIs) are manually extracted by domain experts from the right and left lung areas of CT images. Extracted ROIs are classified to represent 4

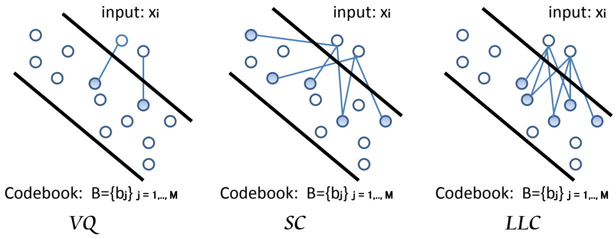

Fig. 3. Comparison among VQ, SC and LLC.The selected codes are highlighted in blue.

states (classes) of emphysema: centrilobular emphysema (CLE), paraseptal emphysema (PSE), interstitial lung diseases (ILD) and normal controls (NL), as shown in Fig. 4 [11]. The representative images were collected from 30 subjects, 9 subjects with CLE, 6 subjects with PSE, 7 subjects with ILD and 8 normal controls (NLs.) Textured tissue samples were selected from 3 representative slices corresponding to the upper, middle and lower lung for each subject. The final datasets we used consisted of 200 lung ROI images with a 1.5 mm and 5 mm spatial resolution. The ROI images are uniformly sized at 100-by-100 pixel, big enough to capture textural differences between 4 tissue classes and small enough to perform analysis efficiently. We classified the CT images based on the classification of the corresponding textures.

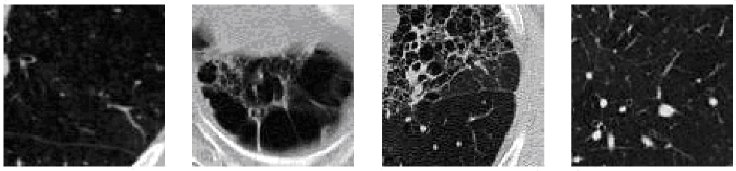

Fig. 4. Texture patterns corresponding to CLE, PSE, ILD and NL classes.

3.2 Methods and Results

The patch size of dense sift feature is 8×8, and the grid size is 4. When generating codebook, we randomly select 400 dense SIFT features from each image to form the feature set. The number of clusters in k-means clustering is set to 450. In SPM, we extract features in 3 layers, each region contain 1, 4, 16 subregions respectively. In FC, we choose 3-by-3 evenly located points as the center points. When conducting approximated LLC, we choose 5 nearest neighbor.

LLC coding is compared with a baseline VQ Coding. We also evaluate the performance of FC and SPM against traditional histogram approach. KNN classifier and linear SVM are used to evaluate our approach. In order to better evaluate the performance of our algorithm, we go through the training and testing phases for 30 times and count the average accuracy and standard deviation in the 30 rounds. At each time, the dataset is randomly divided into training set and testing set. The number of samples in each set is fixed. In this experiment we have two settings for SVM classification: First using 30 for training and 20 for testing, then using 40 for training and 10 for testing in each class. For KNN classification, we use an additional leave-one-out scheme.

As seen in the Table table 2 and Table 1, the proposed approaches outperforms baseline VQ+Histogram method in all settings. Generally, LLC gains higher accuracy rate then VQ. SPM and FC performs better then pure Histogram. SPM produces slightly better result then FC. The highest classification rate of 89.2% is achieved when using SPM+LLC+KNN.

The classifier may fail on some samples because the training set is small and the testing sample doesn't have similar instances in training set. This conjecture is consistent with the fact that we achieve better result when using 40 training samples than when using 30 in SVM experiments, and that KNN performs better in leave-1-out setting, compared to 40 training 10 testing.

	train 30 / test 20	train 40 / test 10
Hist+VQ	64.7% $\pm$ 0.037	65.5% $\pm$ 0.075
FC+VQ	78.6% $\pm$ 0.041	79.0% $\pm$ 0.062
SPM+VQ	78.8% $\pm$ 0.047	81.1% $\pm$ 0.050
Hist+LLC	73.3% $\pm$ 0.036	75.9% $\pm$ 0.052
FC+LLC	79.5% $\pm$ 0.048	83.2% $\pm$ 0.069
SPM+LLC	**80.4%** $\pm$ **0.042**	**83.3%** $\pm$ **0.050**

Table 1. Comparison of classification performance using different methods and parameters. SVM classifier is applied.

	train 40 / test 10	leave-1-out
Hist+VQ	69.5% $\pm$ 0.063	71.3% $\pm$ 0.236
FC+VQ	73.6% $\pm$ 0.039	75.7% $\pm$ 0.112
SPM+VQ	78.7% $\pm$ 0.059	78.2% $\pm$ 0.179
Hist+LLC	76.8% $\pm$ 0.066	77.3% $\pm$ 0.196
FC+LLC	81.8% $\pm$ 0.079	83.5% $\pm$ 0.165
SPM+LLC	**84.0%** $\pm$ **0.069**	**89.2%** $\pm$ **0.169**

Table 2. Comparison of classification performance using different methods and parameters. KNN classifier is applied.

4 Conlcusion

In this paper, we propose a locality-constrained linear coding based method for classifying medical images in a lung image dataset. Different from typical lung image classification methods which utilize vector quantization coding, we introduce a new coding method, Locality-constrained Linear Coding(LLC), and demonstrate its advantages. We also evaluates two kinds of feature representations: spatial pyramid matching and feature context. Experimental result shows that the combined methods achieve state-of-art performance on the lung image dataset. This method can serve as a potential tool for diagnosing lung diseases.

References

1. $http://en.wikipedia.org/wiki/chronic\_obstructive\_pulmonary\_disease$
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. In IEEE Transations on Pattern Analysis and Machine Learning (2002)
3. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. CVPR (2006)
4. Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: Sift flow: dense correspondence across difference scenes. ECCV (2008)
5. Lo, P., Sporring, J., Pedersen, J.J.H., Bruijne, M.D.: Airway tree extraction with locally optimal paths. In MICCAI (2009)
6. Lowe, D.: Object recognition from local scale-invariant features. ICCV (1999)
7. Megalooikonomou, V., Barnathan., M., Kontos, D., Bakic, P., Maidment, A.: A representation and classification scheme for tree-like structures in medical images: Analyzing the branching pattern of ductal trees in x-ray galactograms. IEEE Transactions on Medical Imaging (2009)
8. Mir, A.H., Hanmandlu, M., Tandon, S.N.: Texture analysis of ct images. In IEEE Engineering in Medicine and Biology Magazine (1995)
9. Mller, N.L., Staples, C.A., Miller, R.R., Abboud, R.T.: density mask. an objective method to quantitate emphysema using computed tomography. Chest, vol. 94, no. 4, pp. 782C787, Oct. (1988)
10. Nuzhnaya, T., Barnathan, M., Ling, H., V.Megalooikonomou, Bakic., P., Maidment, A.: Probabilistic branching node detection using adaboost and hybrid local features. Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on (2010)
11. Nuzhnaya, T., Megalooikononou, V., Ling, H., Kohn, M., Steiner, R.: Classification of texture patterns in ct lung imaging. SPIE (2011)
12. Soensen, P.L., Shaker, S.B., de Bruijne, M.: Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Transactions on Medical Imaging, Vol. 29, No. 2, Feb. (2010)
13. Tuceryan, M., Jain, A.K.: Texture analysis. in The Handbook of Pattern Recognition and Computer Vision, 2nd edition (1998)
14. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classfication. CVPR (2010)
15. Wang, X., Latecki, L.J.: Feature context for image classification and object detection. CVPR (2010)
16. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. CVPR (2009)