RESEARCH ARTICLE OPEN ACCESS

An Automatic Classification of Lung Tissue Using Density Based Approach

Mrs. M. Vinthu Kumari[1], Mrs. J. Vijila[2] M. E.
[1]*M.E Final year, University College of Engineering,Nagercoil,Tamilnadu.*
[2]*Assistant Professor, Computer Science & Engineering Department,*
University College of Engineering,Nagercoil,Tamilnadu.
[1]binthukumari@gmail.com
[2]vijilaebenezer@gmail.com

Abstract
The main objective is to develop a technique for classifying the lung tissue as well as finding lung cancer and ranking its severity. To classify the lung tissue the existing system is used. In the existing system Sparse based features are used for classification. In proposed system the lung cancer is detected and segmented. To do this the bee colony technique is used. Before applying the bee colony pre-processing techniques are considered to remove the noise and contrast enhancement. After segmentation the next step is to ranking the severity of the cancer tissues. To perform this process the density based approach is used.
Keywords**—**Gabor LBP,labeling,multi-coordinate HOG, patch adaptive, Rotation Invariant, bee colony.

I. INTRODUCTION

More than 150 disorders of the lung parenchyma is identified as Interstitial lung disease (ILD), which eventually affects breathing. Determining the specific type of disorder is important for treatment. High Resolution Computed Tomography (HRCT) imaging is very useful in classifying the type of tissue. However, manual interpretation of these images is error prone. Hence, an automatic system is needed to differentiate the tissue patterns. Five categories of lung tissues on HRCT images- normal, emphysema, ground glass, fibrosis, and micro-nodules are highly prevalent among the main types of ILDs.

A. Existing Methods

Classification of tissues is performed in two stages: feature extraction and labeling of image categories. Feature extraction is done using filters,and descriptors such as the Scale-Invariant Feature Transform (SIFT), the Local Binary Patterns (LBP) and the Histogram of Oriented Gradients (HOG) is incorporated for lung imaging. These methods produce effective results and even better results by enhancing the descriptor designs based on medical imaging characteristics. Hence patch-based processing is done as it is easier to compute features.

After the extraction of features, labeling is performed for image classification. The most commonly used classifiers include K-Nearest Neighbor (kNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Bayesian classifiers and Artificial Neural Network (ANN). Of these, SVM is highly effective but would be error prone if the feature spaces exhibit considerable overlaps.

B. Proposed Method

A new image classification method for lung tissue patterns, based on feature-based image patch approximation is proposed [1]. Initially, as set of texture, intensity and gradient (TIG) features are extracted for each image patch and two feature descriptors: Rotation invariant Gabor-LBP (RGLBP) for representing rich texture features and Multi-Coordinate HOG (MCHOG) to extract the gradient features are proposed. Then each image patch is classified based on reference dictionaries with a new Patch-Adaptive Sparse Approximation (PASA) algorithm, by which 1) the image patch labeling is enhanced with a statistical measure of the sparse coefficients to measure the minimum discrepancy, 2) a patch-specific adaptation method is designed based on pairwise feature distances to alter the feature values of the reference dictionaries for more discriminative approximation, 3) a feature-space weighting scheme is designed based on overlapping of feature distributions for feature distance computation. Finally, the labeling of the annotated ROI (AROI) is obtained based on probabilistic estimation from the patch-wise classification.

II. TIG FEATURE EXTRACTION

TIG feature extraction process is done as shown in Fig. 1. An image patch is denoted as $p_i : i = 1, \ldots X \times Y\}$, comprising pixels,

and is the intensity value of pixel . The pixel is also indicated by its coordinate as *p(x,y)* and corresponding intensity as *I(x,y)*.The TIG feature set *f(P)* is obtained by combining texture, intensity and gradient features that are extracted for each image patch *P* as discussed below.

A. Texture Description

To incorporate rich texture information while attempting to minimize intra-category variations, we choose to design a new Rotation-Invariant Gabor-LBP (RGLBP) texture descriptor toincorporate the multi-scaleproperty ofGabor

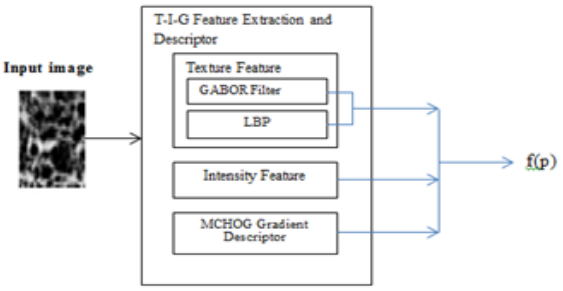

Fig.1 TIG feature extraction

filtersand the rotation-invariant property of LBP features. Let ,) represent the Gabor functions constructed from the basis function *g(x,y)* with *s* = 0,....*S*-1 and *r* = 0,.....*R*-1, where *S* and *R* denote the number of scales and orientations. A set of Gabor-filtered images are then computed by convolving image *I* with each Gabor function. The rotation-invarient Gabor-filtered image is obtained using (1).

$$) = \sum_{r=0}^{R-1} I^{s,r}(x,y) \quad (1)$$

Then rotation-invariant LBP feature $p_i)$ is computed for each pixel as

$$LBP_s(p_i) = min\{ROR(LBP'_s(p_i|8,1), n)\}$$

$$,p_i|8,1) = \sum_n 1(I_i^s > I_{neigh(i,n)}^s)2^n \quad (2)$$

where denotes the intensity of pixel in image ,*neigh(i,n)* indexes the neighboring pixels of the center pixel and *n* = 0,....7. The LBP features $p_i)$ of all pixels in the image patch *P* are accumulated as a histogram feature $P_s(P)$.

$$P(P) = \{RGLBP_s(P): s = 0, \dots, s-1\} \quad (3)$$

B. Intensity Description

Other than the normal intensity information, which is used only for lung CT images that exhibit darker appearances, an intensity histogram *IH(P)* is also used.

C. Gradient Description

Among the various types of gradient-based features, the HOG feature [2] is said to be effective when coupled with LBP features. But it is invariant to rotations. Hence, a new Multi-Coordinate HOG (MCHOG) [3] descriptor is designed.

$$P = \{RS_v(\alpha_{v,1}, \alpha_{v,2}): v = 1, .. V\}$$

$$: \frac{2v\pi}{v} - \frac{\pi}{v} \quad ; \quad : \frac{2v\pi}{V} + \frac{\pi}{V} \quad (4)$$

For each radial section , a coordinate system is defined as

$$\left(1, -\angle \frac{\alpha_{v,1} + \alpha_{v,2}}{2}\right) \quad (5)$$

$$- y_v \quad (6)$$

The gradient orientation and magnitude for each pixel is given by

$$GO_v(p_i) = arctan\left\{\frac{h_{y_v}(p_i)}{h_{x_v}(p_i)}\right\}$$

$$p_i) = \sqrt{h_{y_v}(p_i)^2 + h_{x_v}(p_i)^2} \quad (7)$$

where) and $_i)$ are the gradient values at in and directions. The image patch is divided into *K* overlapping cells $CS_k: k = 0, \dots K-1\}$. Then $OG(CS_k)$ is computed for each cell and patch-level histogram feature *MCHOG(P)* is constructed by concatenating the cell-level histograms.

$$MCHOG(P) = \{MCHOG(CS_{k'})\}$$

where $k_0, \dots, k-1, 0, \dots, k_0 - 1$

$$^{argmax}_{k} \sum_i I_i : \forall P_i \in CS_k. \quad (8)$$

D. Feature Vector

Finally, a patch-wise TIG feature vector *f(P)* is extracted for each image patch *P*

$$= \{RGLBP(P), IH(P), MCHOG(P)\} \quad (9)$$

III. APPROXIMATIVE PATCH CLASSIFICATION

We design a Patch-Adaptive Sparse Approximation (PASA) method, to classify an image patch *P* into five tissue categories as shown in fig. 2.

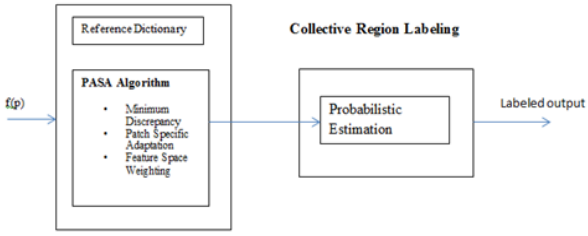

Fig. 2. Classification and labeling

The five tissue categories-normal, emphysema, ground glass, fibrosis and micro-nodule are denoted as $, T_G, T_F$ and T_M and the objective is to assign a label $\in \{T_N, T_E, T_G, T_F, T_M\}$.

A. Minimum Discrepancy Criterion

Assume that there are *Q* image patches belonging to a certain tissue category in the training set. An over-complete feature dictionary matrix for *Q* reference image patches is then constructed by concatenating their features as column vectors.

$$\{f(P_q): q = 1, \dots, Q\} \in R^{HXQ} \quad (10)$$

where*H* is the feature dimension and *H* < *Q*. Five dictionaries are created with one for each tissue category. To classify an image patch *P*, a sparse-regularized linear model is formulated to compute an approximated vector of its feature *f(P)* from the feature dictionary .

$$\underset{w_l}{argmin} \|f(P) - D_l w_l\|_2^2 \; s.t. \|w_l\|_0 \leq C \quad (11)$$

$$= D_l w_l \quad (12)$$

where is sparse coefficient vector and *C* is a constant. After optimizing using Orthogonal Matching Pursuit(OMP) algorithm[4], labeling *L(P)* of image patch *P* is obtained as

$$= \underset{l}{argmin} \|f(P) - f_l'(P)\|_2 \sigma(w_l) \quad (13)$$

where is the standard deviation which is used to improve labeling efficiency.

B. Patch-Specific Adaptation

To obtain accurate labeling, assuming an image patch *P* is of category, we would expect that its feature approximation based on dictionaries $= T_N$ would produce larger discrepancies with *f(P)* than $= T_N$ and hence labeling of *P* would be . This patch-specific adaptation is helpful to achieve discriminative feature approximation. Initially a feature distance $_q$) between image patch *P* and reference patch is computed using histogram-intersection distance rather than Euclidean distance to reduce the misrepresentation of feature similarities

$$_q) = \lambda . \left\{ \frac{|f(P) - f(P_q)|}{f(P) + f(P_q)} \right\} \quad (14)$$

where is a vector of feature weights. A scaling coefficient $_q)$ is derived to rescale the feature vector into) and the design of $_q)$ is based on the degree of similarity between *P*and , and the sign of distance between *f(P)* and *f( ).*

$$_) = \begin{cases} \exp(-d(P, P_q)) & \text{if } f(P) \geq f(P_q) \\ 2 - \exp(-d(P, P_q)) & \text{otherwise} \end{cases} \quad (15)$$

Then the feature vector is rescaled into vector) by

$$) = \phi(P, P_q) f(P_q) \quad (16)$$

Finally, the scaled feature vectors) are assembled as the transformed reference dictionary as

$$\{f^{\phi}(P_q): q = 1, \dots Q\} \in R^{H \times Q} \quad (17)$$

and integrated into sparse linear model as

$$\underset{w_l}{argmin} \|f(P) - D_l^{\phi} w_l\|_2^2 \; s.t. \|w_l\|_0 \leq C \quad (18)$$

C. Feature-Space Weighting

Different feature dimensions should contribute differently to distance computation

between feature vectors, and hence naturally feature weights are desired. First, we denote a feature element in the TIG feature vector as and $= \{f_{v_{q,h}} : h = 1, \dots H\}$. Then the dictionary is denoted as concatenation of row vectors

$$D_l = \{fr_h : h = 1, \dots H\} \in R^{H \times Q}$$

$$\{fv_{1,h}, fv_{2,h}, \dots, fv_{Q,h}\} \in R^{1 \times Q} \quad (19)$$

Then for a certain feature dimension *h*, a probability distribution $D_l)$ of the reference patches is created using the kernel density estimation.

$$D_l, \mu_h) = Q^{-1} \sum_q \kappa(\mu_h - fv_{q,h}) \quad (20)$$

where takes a value in the feature range of

$$\left[\min_q fv_{q,h}, \max_q fv_{q,h}\right] \quad (21)$$

and *(.)* is the Gaussian kernel. A set of five probability distributions are created for a feature dimension $_h(D_l)\}$, and a degree of feature overlapping is computed as

$$= \frac{\sum_{\mu h} [\sum_l PD_h(D_t, \mu_h) \wedge tPD_h(D_t, \mu_h)]}{\sum_{\mu h} \sum_l PD_h(D_l, \mu_h)} \quad (22)$$

Finally the feature weight is computed as

$$\lambda_h = \{1 + exp[-2(1 - FO_h - |FO|)]\}^{-1}$$

$$= H^{-1} \sum_h (1 - FO_h) \quad (23)$$

D. Dictionary Construction

To construct the reference dictionary, the whole database is first divided sequentially into three sets of roughly equal numbers of 3-D images. The patch-wise feature vectors)} from all of the reference images are concatenated together according to their ground truth labeling to construct five feature dictionaries $= T_N, T_E, T_G, T_F, T_M\}$.

IV. COLLECTIVE REGION LABELING

A region-level classification is performed to achieve a unanimous label for each AROI, based on collective probabilistic estimation of its image patches. A single label for the AROI $\mathcal{I}) \in \{T_N, T_E, T_G, T_F, T_M\}$ is obtained. First, rather than using discrete labeling , five probability values are computed for each image patch . The probability value $_l, l)$ is derived based on the discrepancy between its feature vector and the approximation

$$_l, l) = exp\left(-\frac{2\|f(P_\alpha) - f_l'(P_\alpha)\|_2 \sigma(w_l)}{\sum_{l'} \|f(P_\alpha) - f_{l'}'(P_\alpha)\|_2 \sigma(w_{l'})}\right) \quad (24)$$

Hence the final labeling is given as

$$\mathcal{I}) = \arg\max_l \sum_\alpha PR(P_\alpha, l) \quad (25)$$

V. CONCLUSION AND FUTUE WORK

An automatic classification method for lung HRCT images is proposed. The challenges in low inter-class distinctions and high intra-class variations are tackled by designing a feature-based image patch approximation method. First, an image patch is represented as a feature vector, based on RGLBP texture and MCHOG gradient descriptors, and then classified into one of the five tissue categories using PASA classifier. Finally, a single labeling is assigned for each AROI based on collective probabilistic estimation. Before applying the bee colony pre-processing techniques are considered to remove the noise and contrast enhancement. After segmentation the next step is to ranking the severity of the cancer tissues. To perform this process the density based approach is used.

REFERENCES

[1] Feature-Based Image Patch Approximation for Lung Tissue Classification, Yang Song, Student Member, IEEE, WeidongCai, Member, IEEE, Yun Zhou, and David Dagan Feng, Fellow, IEEE ,IEEE Transa.on Medical Imaging, Vol. 32, No. 4, April 2013.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. Comput. Vis. Pattern Recognit., 2005, pp. 886–893.

[3] B. Fan, F. Wu, and Z. Hu, "Aggregating gradient distribution into intensity orders: A novel local image descriptor," in Proc. Comput. Vis.PatternRecognit., 2011, pp. 2377–2384.

[4] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation,"*IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct.2004.