TEXTURE CLASSIFICATION OF SARS INFECTED REGION IN RADIOGRAPHIC IMAGE

Xiaoou Tang[1], Dacheng Tao[1], and Gregory E. Antonio[2]

[1]Department of Information Engineering,
The Chinese University of Hong Kong, Shatin, Hong Kong
{xtang, dctao2}@ie.cuhk.edu.hk
[2]Department of Diagnostic Radiology and Organ Imaging
Prince of Wales Hospital,
The Chinese University of Hong Kong, Shatin, Hong Kong
gregantonio@cuhk.edu.hk

ABSTRACT

In this paper, we conduct the first study on SARS radiographic image processing. In order to distinguish SARS infected regions from normal lung regions using texture features, we propose several improvements to the traditional gray-level co-occurrence texture features [2]. We use a multi-level feature selection approach to extract texture features from a multi-resolution region based co-occurrence matrix directly for texture classification. The selected texture features can preserve most of the discriminant information in the texture image. Satisfactory results are obtained on a large set of chest radiographic images of SARS patients.

1. INTRODUCTION

Severe Acute Respiratory Syndrome (SARS) outbreak in Hong Kong started in March 2003 and quickly spread to many regions around the world. By the end of the epidemic, there were 1,755 patients infected and 299 deaths in Hong Kong [1]. The main symptoms of SARS are high fever and dry cough, shortness of breath or breathing difficulties. SARS may also be associated with other symptoms including a headache. Because of the highly contagious nature of the disease and its very fast progress that often threatens the life of the patient, it is critically important to identify the disease at an early stage. However, since most of the symptoms are similar to regular pneumonia and fever, it is very difficult to give an accurate diagnosis of the disease. All currently available methods depend on laboratory testing of the virus samples from the patient, which is both costly and time consuming.

In this paper, we study the chest radiographs of the SARS patients to investigate a possible computer-aided approach to distinguish the SARS infected area from the normal lung area. This can be an important first step toward image based computer-aided diagnosis. Of course, it is unrealistic to expect accurate diagnosis only based on automatic computer processing of radiographic images. However, we do expect our study to be able to assist doctors with their diagnosis in the future. In addition, since for confirmed patients the chest radiographic images are taken everyday, we can also compare the progress of the images with previous patients in the database to monitor the effect of the treatment.

Because SARS regions are irregular, we cannot use shape to distinguish it from normal areas. So we focus on using texture classification to classify the SARS region. In this paper, we propose several improvements to the classic texture model, gray-level co-occurrence matrix [2], to distinguish the subtle SARS texture. We use a multi-level feature selection approach to extract texture features from a multi-resolution region based co-occurrence matrix directly for texture classification. Encouraging results are obtained on a set of chest radiographic images.

2. MULTI-RESOLUTION REGION BASED TEXTURE FEATURE

2.1. Region Based Co-occurrence Matrix

Co-occurrence texture features were proposed by Haralick et al. [2]. For an image with N by N pixels and G gray levels, the co-occurrence matrix for a displacement d in a direction q is defined to be a G by G matrix whose entry M(i, j) is the number of occurrences of transitions from gray level i to gray level j, given the inter-sample distance d and the direction q. The matrix gives a measure of the joint probability density of the pairs of gray levels

0-7803-8554-3/04/$20.00 ©2004 IEEE.

that occur at pairs of points separated by distance *d* in the direction *q*. For a coarse texture, *d* is relatively small compared to the sizes of the texture elements; the pairs of points at separation *d* have similar intensity values. This means the matrix M has large values near its main diagonal. Conversely, for a fine texture the values in M are quite uniformly spaced. Thus, a measure of the degree of value spread around the main diagonal of M should provide a good sense of the texture coarseness. Similarly, one can extract other features to measure the directional information, contrast, correlation, etc. Haralick et al. [2] proposed 28 second-order statistic features that can be measured from this co-occurrence matrix.

Generally, the co-occurrence matrix is computed from a rectangular region or image. In our application, however, the regions are not rectangles. In order to compute the texture features, we develop a region based co-occurrence matrix:

1. Extract the marked SARS infected regions.
2. Find the maximum bounding box of each region.
3. Quantize the region with a given bin number and fill the blank part of the bounding box with –1.
4. Calculate the co-occurrence matrix $\mathbf{P}$ of the filled bounding box, and extract a sub-matrix $\mathbf{P}_s$ from $\mathbf{P}$, where $\mathbf{P}_s$ is obtained by deleting the first row and the first column of $\mathbf{P}$. The region based co-occurrence matrix is:

$$\mathbf{P}_s = \mathbf{P}_{2:n,2:n} \tag{1}$$

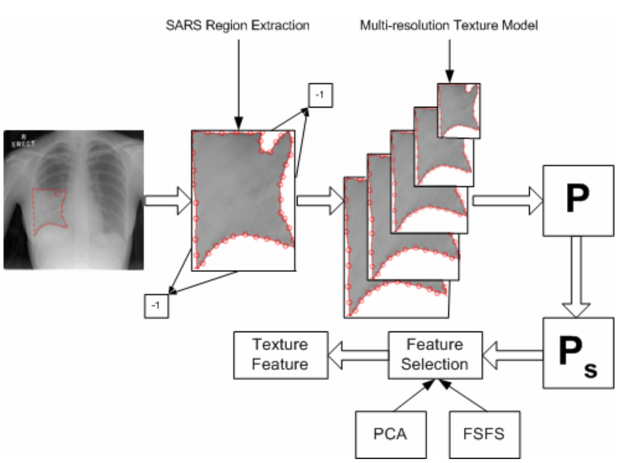

Figure 1. Flowchart of the multi-resolution non-rectangle region's co-occurrence matrix and texture feature extraction.

The size of the texture image is crucial for classification. To preserve more information of the texture image, we select a series scales to zoom the original image and the scale vector is: $S = [0.1\ 0.2\ 0.4\ 0.6\ 0.8\ 1.0]$, where $S = 0.6$ means the ratio between the size of the zoomed image and the original image is 0.6. For each scale, the region-based co-occurrence matrix and the corresponding statistical features are calculated.

2.2. Multi-level Feature Selection

The original texture features computed from the co-occurrence matrix are mostly based on intuitive observation of the shape and statistics of the matrix [2]. There are two drawbacks with this approach. First, there is no theoretical proof that, given a certain number of features, maximum texture information can be extracted from the co-occurrence matrix. Second, many of these features are highly correlated with each other. A better approach is to use the co-occurrence matrix as the texture feature vector directly to preserve all the information in the matrix instead of developing new functions to extract texture information. However, this again introduces two problems: the large dimensionality of the feature vector and the high-degree correlation of the neighborhood features. To alleviate these problems, we developed a multi-level dominant eigenvector estimation (MDEE) method to approximate PCA and apply to the co-occurrence matrix directly to extract texture features [3].

The MDEE cuts a long feature vector into sections of small vectors, and then performs a PCA on each small vector separately. The selected top features with large eigenvalues in each section are then combined to form a new feature vector with a second PCA applied again. Several orders of computation complexity reduction from the conventional PCA are achieved by this method.

In [4], a new feature similarity based feature selection (FSFS) method is developed and shown to perform better than PCA for feature selection. In this paper, we select the maximal information compression index (λ_2) as the feature similarity measure. Let Σ be the covariance matrix of random variables x and y. Define maximal information compression index as $\lambda_2(x,y) =$ smallest eigenvalue of Σ, i.e.,

$$2\lambda_2 = \left(\operatorname{var}(x) + \operatorname{var}(y)\right) - \sqrt{\left(\operatorname{var}(x)+\operatorname{var}(y)\right)^2 - 4\operatorname{var}(x)\operatorname{var}(y)\left(1-\rho(x,y)^2\right)}. \tag{2}$$

The larger the value of λ_2, the less of the dependency of the two variables. The value of λ_2 is zero means the features are linearly dependent. For feature selection, FSFS first partitions the original feature set into a number of homogeneous subsets and select a representative feature from each subset based on the similarity measure.

However, the FSFS method encounters the same problem as PCA. The computational complexity of FSFS is $\boldsymbol{O}(D^2 l)$, where D is the feature dimension and l is the size of the data-set. In our study, the feature dimension is 1024×6. The computational cost is too high for FSFS. In order to overcome this problem, we propose a similar multi-level approach as the MDEE method. We first apply the FSFS to feature vector for each image scale, then combine the selected features and use the FSFS again on the combined feature vector.

The flowchart of our feature selection algorithm is shown in Figure 2. For each level, we calculate the region-based co-occurrence matrix, and then FSFS or PCA is applied to the matrix directly to select first level features. We then combine all the selected features into a new feature vector and the feature selection method is used again to select the final features.

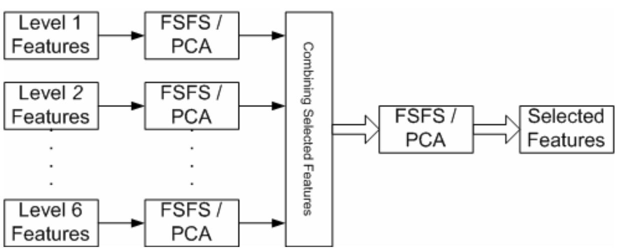

Figure 2. Flowchart of the multi-level feature selection method.

3. EXPERIMENTAL RESULTS

In this Section, we use the new algorithm to classify the SARS infected region from the normal lung region in chest radiographic images. We also compare the new features with traditional co-occurrence features. We use the support vector machine (SVM) [5] with Gaussian Kernel as the classifier since SVM is a very effective binary classifier. All the parameters are default values in OSUSVM [6].

3.1. Data Set

We use the posteroanterior chest radiographs taken by the department of Diagnostic Radiology & Organ Imaging of the Prince of Wales Hospital. The digital images were obtained by digitizing the chest radiographs in the SIEMENS medical computer system. The original image has a pixel size of about 0.175mm, a matrix size of about 2000×2400, and a gray level range of 16 bits. The SARS infected regions and normal regions of all lung radiographs were labeled by doctors in the hospital as ground truth. Table 1 shows the details of the database and Figure 3 shows some sample images and SARS infected regions in the database.

Table 1. Image Database

	SARS	Normal
Training	37	37
Testing	126	38
Total	163	75

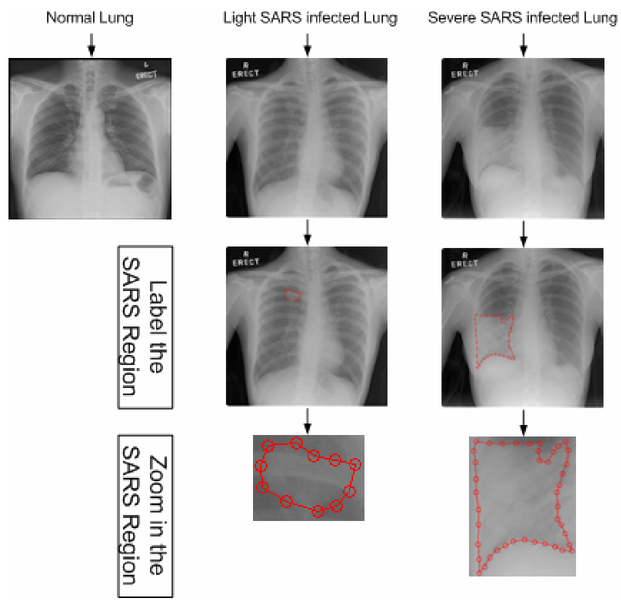

Figure 3. Sample images and SARS infected regions in the database.

3.2. Classification Using Traditional Features

We first use the traditional texture features defined in [2] to classify the images. Classification results are summarized in Table 2. The results show that the traditional feature of each co-occurrence direction in each scale cannot discriminant the SARS and normal lung regions well. Figure 4 shows the results of using FSFS to select features from all the traditional texture features (the original feature dimension is $13 \times 6 \times 4 = 312$). The classification result is still less than satisfactory.

Table 2. Traditional feature based classification. The first row is the scale value of the image and the first column is the direction of the co-occurrence matrix.

	0.1	0.2	0.4	0.6	0.8	1.0
0	0.7927	0.8110	0.8110	0.2561	0.8110	0.8171
45	0.7683	0.2500	0.2500	0.8171	0.7927	0.8171
90	0.7683	0.2561	0.8354	0.2561	0.8232	0.2561
135	0.2439	0.2317	0.2439	0.8171	0.8232	0.8476

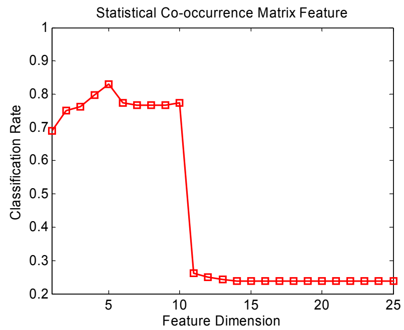

Figure 4. Classification results using FSFS to combine traditional co-occurrence features.

3.3. Classification Using the New Features

Classification results using the multilevel PCA to extract texture features directly from the co-occurrence matrices are shown in Figure 5. The recognition rate is significantly improved over the traditional features. This shows that the method can effectively preserve the discriminant texture information for SARS and normal lung region classification.

Next, we use the new multi-level FSFS method to extract the texture features directly from the original co-occurrence matrix. The recognition rate is further improved as shown in Figure 6. The highest classification rate of the method is around 97%.

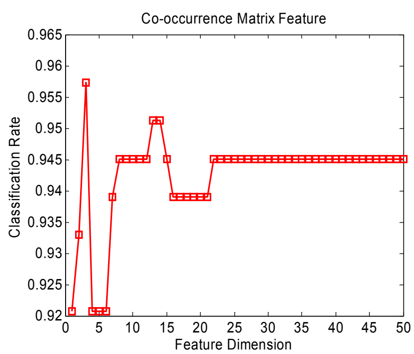

Figure 5. Classification results using multi-level PCA to extract texture features directly from the co-occurrence matrices.

4. CONCLUSION

Toward assisting doctors to diagnose SARS patients, we conduct a preliminary study on texture classification of SARS infected regions in chest radiographic images. In order to distinguish SARS infected regions from normal lung regions, we propose several improvements to the traditional gray-level co-occurrence texture features. We use a multi-level feature selection approach to extract texture features from a multi-resolution region based co-occurrence matrix directly for texture classification. The multi-level Feature Similarity-based Feature Selection algorithm is shown to be very effective in preserving most of the discriminant information in the texture image. Experiments on a large set of chest radiographic images of SARS patients show encouraging results. This is a promising first step toward computer-aided diagnosis of the disease.

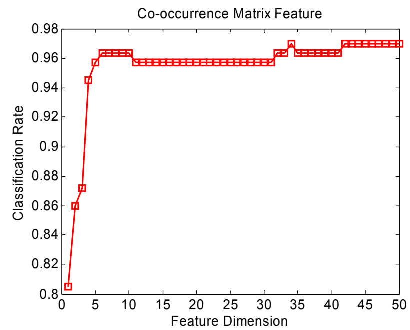

Figure 6. Classification results using multi-level FSFS to extract texture features directly from the co-occurrence matrices.

5. ACKNOWLEDGMENTS

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong SAR (Project no. AOE/E-01/99).

6. REFERENCES

[1] K. T. Wong, G. E. Antonio, D. S. Hui et. al., "Severe acute respiratory syndrome: radiographic appearances and pattern of progression in 138 patients," Radiology, 228(2): 401-406, Aug. 2003.

[2] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. on Systems man and Cybernetics, 3, pp 610-621, 1973.

[3] X. Tang, "Texture information in run-length matrices," IEEE Transactions on Image Processing, vol. 7, No. 11, pp. 1602-1609, Nov. 1998.

[4] M. Pabitra, C. A. Murthy, and K. Sankar, "Unsupervised Feature Selection Using Feature Similarity," IEEE Trans. on Pattern Analysis and Machine Intelligence, 24, pp. 1-13, 2002.

[5] V. Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, New York (1995).

[6] http://www.eleceng.ohio-state.edu/~maj/osu_svm/.