ADJUSTED POLYNOMIAL FEATURES FOR ANALYSIS OF LUNG CT IMAGES

A.V. Gaidel

Samara National Research University, Samara, Russia,
Image Processing Systems Institute - Branch of the Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Samara, Russia

Abstract. We introduced new polynomial features described as polynomials on the image pixels domain. After imposing natural conditions these features transform into linear combinations of the image autocovariance function readings. We proposed a way to adjust these features using learning sample textural properties. Experiments on a real diagnostic dataset of lung CT images showed a decrease of the error probability in comparison with the previous works.

Introduction

Computer-aided diagnostics provides automatization, objectivization and cheapening of medical procedures. It also belongs to the diagnostic image analysis performed by medical professionals using the lesion detection. Emphysema diagnostics by 2D slices of lung CT images is an example of such procedure (Fig. 1). In the presence of the disease one can note specific dark areas on the image.

One of the most important and least studied steps of the image recognition process is the selection of the features, which quantitatively describes the characteristics that are most different on the images of different classes. Development of the automatic feature construction methods appropriated to the particular task is an open problem.

For the selection of a small finite number of features one can use the brute force, allowing him to select the best subset of features for specific applications [1]. If the number of features exceeds several tens, heuristic selection algorithms are used, one of which was previously used to analyze lung CT images [2]. In addition, even early texture analysis was used to analyze these images in [3] mostly with a help of run-length features.
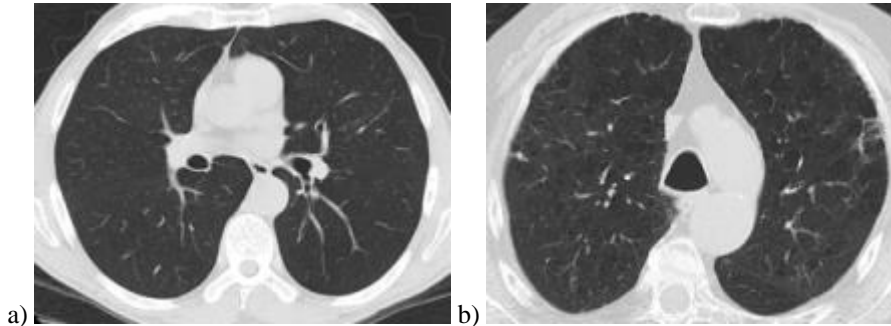
Fig. 1. Lung CT images: healthy (a), affected by emphysema (b)

Approaches to the selection of the infinite collections of features are much less studied. In most papers, such as [4] and [5], the new features are selected as a number of several sequential transforms of the existing primitive features. However, approaches to the selection of certain feature parameters using evolutionary algorithms are also proposed [6]. The theory of the efficient linear local features construction is described in [7] and [8].

The general approach to the adjusted feature selection is described in [9]. There was introduced a simple adjusting algorithm based on image rotation.

The aim of this work was to use the new adjusted polynomial features for the lung CT images classification, as well as the study of the feature space quality criteria and adjusting algorithms, the most suitable for this problem.

Polynomial features and quadratic features

Let $\mathbf{\Omega}$ be a set of images we want to recognize, and each image is as follows

$$\omega(m,n):D_{\omega}\to\mathbf{Q},$$

where the pixels domain $D_{\omega}\subseteq\mathbf{Z}^2$ is a region of interest, and $\mathbf{Q}=[0,Q-1]\cap\mathbf{Z}$ is a set consisting of Q gray levels.

We constructed polynomial features as the polynomials on $\mathbf{\Omega}$. Say that a multi-index of degree $q\in\square_0$ is a tuple

$$\gamma=\begin{pmatrix}\gamma(1) & \gamma(2) & \ldots & \gamma(|D_{\omega}|)\end{pmatrix}^T$$

such that $\forall k\in\left[1;|D_{\omega}|\right]\cap\mathbf{Z}:\gamma(k)\in\square_0$ and

$$\sum_{k=1}^{|D_{\omega}|}\gamma(k)=q.$$

Here we have $\square_0 = \left\{ j \in \mathbf{Z} \mid j \geq 0 \right\}$, and the operator $\left| A \right|$ denotes a number of elements in the finite set A.

Also suppose there is a lexicographical order on D_ω such as

$$\left(\left(m_1, n_1 \right) \prec \left(m_2, n_2 \right) \right) = \left(m_1 < m_2 \vee m_1 = m_2 \wedge n_1 \leq n_2 \right)$$

and all pixels $(m,n) \in D_\omega$ are numerated in correspondence with that order. It defines a finite sequence of pixels $\left\{ \left(m_k, n_k \right) \right\}_{k=1}^{\left| D_\omega \right|}$ for which

$$\forall i, j \in \left\{ 1, 2, \ldots, \left| D_\omega \right| \right\} : (i < j) \Rightarrow \left(\left(m_i, n_i \right) \prec \left(m_j, n_j \right) \right).$$

Assuming $\mathbf{I}_q$ is a set of all multi-indices of degree q, we can introduce a polynomial of degree q as

$$\Psi_P (\omega, \theta) = \sum_{p=0}^{q} \sum_{\gamma \in \mathbf{I}_p} \theta(\gamma) \prod_{k=1}^{\left| D_\omega \right|} \left(\omega \left(m_k, n_k \right) \right)^{\gamma(k)}, \tag{1}$$

where $\theta(\gamma) : \mathbf{I}_q \to \square$ is its coefficient corresponding to the term with the multi-index γ.

The operator (1) determines the broad collection of parametric features differing by their coefficients θ, some of which may be better or worse appropriated to the particular application. The adjusted polynomial feature selection is to select the best coefficients θ for the specified problem, for which images from the learning sample are well divided into classes in the feature space formed by the operator (1).

In [10] we proved that under natural constraints related to the invariance to some texture-independent transforms polynomial features (1) become the form

$$\Psi_{P(2)} (\omega, \theta) = \sum_{(\Delta m, \Delta n) \in W_d} \theta(\Delta m, \Delta n) \hat{R}_\omega (\Delta m, \Delta n), \tag{2}$$

where $\hat{R}_\omega (\Delta m, \Delta n)$ are the readings of the image autocovariance function

$$\hat{R}_\omega (\Delta m, \Delta n) = \frac{1}{\left| D_\omega (\Delta m, \Delta n) \right|} \sum_{(m,n) \in D_\omega (\Delta m, \Delta n)} \omega(m,n) \omega(m + \Delta m, n + \Delta n)$$

calculated within the square window of radius d

$$W_d = \left[-d; +d \right]^2 \cap \mathbf{Z}^2. \tag{3}$$

Here $D_\omega (\Delta m, \Delta n)$ is understood as a set of pixels (m,n) from the region of interest for which the pixels $(m + \Delta m, n + \Delta n)$ also belong to the region of interest:

$$D_{\omega}\left(\Delta m, \Delta n\right) = \left\{\left(m,n\right) \in D_{\omega} \mid \left(m + \Delta m, n + \Delta n\right) \in D_{\omega}\right\}.$$

In fact, the operator (2) is a linear combination of image autocovariance function readings. This is only a subset of quadratic polynomials.

We assumed that the images from the learning sample are standardized, i.e. have zero mean value and unit standard deviation. Standardization consists of the mean value subtraction and the dividing by the standard deviation.

Feature space quality criteria

In this work we used the operator (2) to calculate features and the nearest neighbor algorithm as a classifier, but we calculated a distance between feature vectors by their standard score. We determined the coefficients $\theta\left(\Delta m, \Delta n\right)$ for the features (2) solving the optimization problem, where one of the feature space quality criteria computed from the learning sample is used as an objective function. We investigated the following feature space quality criteria.

1. The accuracy of recognition $\tilde{J}$ is a fraction of correctly recognized images from the learning sample calculated using leave-one-out cross-validation [11].
2. Bhattacharya distance

$$\mu\left(\frac{1}{2}\right) = \frac{1}{8}\left(a_1 - a_2\right)^T \left(\frac{R_1 + R_2}{2}\right)^{-1}\left(a_1 - a_2\right) + \frac{1}{2}\ln\left(\frac{1}{2}\frac{\left|R_1 + R_2\right|}{\sqrt{\left|R_1\right|\left|R_2\right|}}\right),$$

where a_l and R_l are the means and the correlation matrices of the feature vectors from the l-th class.

3. The first discriminant analysis criterion from [11]:

$$J_1 = \mathrm{tr}\left(R_{\Sigma}^{-1}R\right),$$

where R_{Σ} is a correlation matrix of feature vectors from the both class distributions mixture and $R = \left(R_1 + R_2\right)/2$ is a mean intraclass correlation matrix.

4. The fourth discriminant analysis criterion from [11]:

$$J_4 = \mathrm{tr}^{-1}\left(R_{\Sigma}\right)\mathrm{tr}\left(R\right).$$

Experiments

Experiments were carried out according to the scheme described in detail in [9]. The dataset for the experiments consists of 160 lung CT images with resolution of

140×200 pixels. Each experiment was carried out separately for each of the quality criteria described in the previous section, and for each of the optimization algorithms suitable for this criterion.

For all criteria except J_1 we used one of three optimization algorithms: simple random search, genetic algorithm or simulated annealing. For the criterion J_1 there is a known optimization algorithm used in the discriminant analysis and based on the principal component analysis [11].

We carried out each set of experiments as follows.

1. We randomly divided the entire set of images into a learning sample and a test sample of equal volumes.
2. With a help of a given optimization procedure we determined coefficients $\theta(\Delta m, \Delta n)$ maximizing the specified quality criterion. The output was a taught-in system that stores in memory all adjusted feature vectors corresponding to the images from the learning sample.
3. Among the adjusted features we selected a subset of the most efficient features using the heuristic feature ordering procedure described in [2].
4. We verified quality of the recognition system by the recognition error rate assessment using the test sample.

Specific parameters of the algorithms described in detail in [9]:

- $d = 2$ is a window (3) range, in which covariance function readings where computed;
- $K = 13$ is a number of the quadratic features;
- $N_{\text{opt}} = 200$ is a number of steps of the global optimization iterative algorithms;
- $N_{\text{pop}} = 10$ is a population size of the genetic algorithm;
- $p_{\text{mut}} = 0,1$ is a mutation probability of the genetic algorithm;
- $t_0 = 10$ is an initial temperature of the simulated annealing;
- $-1 \le \theta(\Delta m, \Delta n) \le +1$ – constraints on the coefficients for the cases in which optimization procedures can't work without constraints.

Table 1. The experimental results of adjusted quadratic features research

Adjusting algorithm	Best criterion	Error	Number of features
Random search	$\mu(1/2)$	0.06	6
Genetic algorithm	J_4	0.12	4
Simulated annealing	J_4	0.15	7
Principal components analysis	J_1	**0.20**	13

Research findings on adjusted quadratic features (2) effectiveness for the problem of automatic emphysema diagnosis by the lung CT images are shown in the Tab. 1. For each adjusting method we presented the best feature space quality criterion, the best error rate of the test sample images, as well as the number of features in the most efficient set. We highlighted the line with the lowest error rate.

Agresti-Coull interval for the best recognition error rate with the 95 % confidence level is (0.02, 0.14). We obtained the best results using random search, but they much surpass their counterparts obtained in [2] and [3] for the same problem.

Conclusion

We investigated new polynomial features allowing adjusting for the textural properties of the learning sample images in the lung CT image recognition problem. Analyzing experimental results we found out the decrease of the recognition error rate from 0.11 to 0.06 in comparison with the results given in [2] and [3]. In the future we plan to investigate the effectiveness of these features to other image analysis problems.

Acknowledgements

References

1. Ilyasova NYu, Kupriyanov AV, Paringer RA. Formation of features for improving the quality of medical diagnosis based on discriminant analysis methods. Computer Optics, 2014; 38(4): 851-855. [In Russian]
2. Gaidel AV, Zelter PM, Kapishnikov AV, Khramov AG. Computed tomography texture analysis possibilities of the chronic obstructive pulmonary disease diagnosis. Computer Optics, 2014; 38(4): 843-850. [In Russian]
3. Ginsburg SB, Lynch DA, Bowler RP, Schroeder JD. Automated Texture-based Quantification of Centrilobular Nodularity and Centrilobular Emphysema in Chest CT Images. Acad Radiol, 2012; 19(10): 1241-151.
4. Neshatian K, Zhang M, Johnston M. Feature construction and dimension reduction using genetic programming. Lecture Notes in Comput. Sc., 2007; 4830: 160-170.
5. Fan W, Zhong E, Peng J, Verscheure O, Zhang K, Ren J, Yan R, Yang Q. Generalized and heuristic-free feature construction for improved accuracy. Proceedings of the 10th SIAM International Conference on Data Mining; Columbus, OH, United States, 29 April – 1 May, 2010: 629-640.
6. Lillywhite K, Lee D-J, Tippetts B, Archibald J. A feature construction method for general object recognition. Pattern Recogn, 2013; 71(3): 514-527.

7. Myasnikov VV. Constructing efficient linear local features in image processing and analysis problems. Automat Rem Contr+, 2010; 31(4): 58-76.
8. Myasnikov VV, Bavrina AY, Titova OA. Analysis of methods for construction of efficient linear local features for digital signals and images description. Computer Optics, 2010; 34(3): 374-381. [In Russian]
9. Gaidel AV. A method for adjusting directed texture features in biomedical image analysis problems. Computer Optics, 2015; 39(2): 287-293 [In Russian]. DOI: 10.18287/0134-2452-2015-39-2-287-293
10. Gaidel AV. Matched polynomial features for the analysis of grayscale biomedical images. Computer Optics, 2016; 40(2): 232-239 [In Russian]. DOI: 10.18287/2412-6179-2016-40-2-232-239.
11. Fukunaga K. Introduction to statistical pattern recognition. San Diego: Academic Press; 1990.