Polynomial feature engineering for classification of textural images

A V Gaidel1,2
1Samara National Research University, Moskovskoe Shosse 34, Samara, Russia, 443086
2Image Processing Systems Institute - Branch of the Federal Scientific Research Centre “Crystallography and Photonics” of Russian Academy of Sciences, Molodogvardeyskaya str. 151, Samara, Russia, 443001

e-mail: andrey.gaidel@gmail.com

Abstract. I consider a number of methods of automatic quadratic features adjustment for digital textural images of biological tissues in order to improve the quality of classification. The proposed approaches are based on optimization procedures that use various quality criteria of a feature space as target functions. I investigate the methods based on random search, genetic algorithm, simulation of annealing, as well as the original hybrid algorithm. I presented results of experimental studies of the proposed algorithms on sets of real X-ray images of bone tissue and the lung CT images. We show that the hybrid algorithm provides more stable results regardless of the chosen quality criterion of the feature space, which is expressed in decreasing of the average percentage of incorrectly recognized images in comparison with the use of the specific optimization methods individually.

1. Introduction

The problem of automatic classification of images occurs in a variety of areas: from medical diagnosis to geoinformatics. Software solutions to this problem are used both in computer vision systems and in decision support systems of various kinds. Moreover, many images analyzed in practice are textural, that is, they do not represent any objects, but carry information in themselves due to gray level differences [1].

There are many solutions to this problem, very different from each other. Recently, the popular approach is the use of convolutional neural networks, some of which were first described in the last century [2]. Recent studies have shown that this approach is extremely effective for a wide class of problems, so that the theory of deep learning has greatly expanded and acquired many details [3]. At the same time, the main disadvantage of convolutional neural networks is still the high computational complexity of their learning process, which does not allow to properly study the efficiency of various configurations of neurons and their other parameters [4].

A more classical approach is that a certain set of features (numerical characteristics of this image) is computed from a given image at first, and then some sorting procedure is applied to the received feature vector. In this case, matrices of mutual occurrence, Gabor filters, Laws' masks, wavelet
transforms, parameters of models of Markov fields and many other things are often used to calculate the features of the textural image [1]. Popular approaches for the classification of feature vectors are the k nearest neighbors algorithm and the support vector machine [5]. Neural networks can also be used for this, as well as for the classification of images as a whole [6].

When using features for the classification of images, there arises the problem of selecting a set of features that describes the images for a particular problem in the best way. To this end, various heuristic procedures can be used, consistently increasing and decreasing the current set of the most effective features [7]. Also, methods that perform an optimal linear transformation of the original feature space are often used to reduce its dimensionality, which reduces the amount of data used to describe images, and also increases the efficiency of classification [8].

In [9], new quadratic features of monochrome texture images were proposed, differing in the possibility of adjusting using the learning sample to improve the efficiency of solving a specific classification problem. In the same paper, some procedures for their adjustment have been described and investigated. These features have been successfully used to solve the applied problem of automated diagnosis of emphysema from images of computed tomography of the lungs [10]. For the X-ray images of femoral neck some other special statistical features were described in [11].

In this paper, we developed the idea of using optimization procedures for automatic adjustment of the parameters of consistent quadratic features. For this purpose, a number of optimization methods and fitness functions are considered in the form of criteria for the quality of the feature space. A new hybrid optimization method is also proposed and experimentally investigated for matching quadratic features. In experimental studies, three sets of data were used: photos of rice, X-Ray images of the femoral neck and the CT images of the lungs.

2. Methods for polynomial features adjustment

2.1. Adjusted quadratic features

Let \( \phi(m, n) \) be a grayscale digital image representing a function that maps pixels from the Region of Interest \( D_\omega \subset \mathbb{Z}^2 \) to the set of gray levels \( Q \subset [0; Q-1] \cap \mathbb{Z} \) which includes \( Q \) gray levels. Here \( \mathbb{Z} \) is a set of integers.

Polynomial features are constructed as polynomials on a set of image pixels. We introduce the multi-index [12] of order \( [1; q] = \gamma \in \mathbb{Z} \) as a vector \( \gamma = (\gamma(1), \gamma(2), \ldots, \gamma(|D_\omega|)) \) for which

\[
\forall k \in [1; |D_\omega|] \cap \mathbb{Z}: \gamma(k) \in [0; +\infty) \cap \mathbb{Z} \quad \text{and} \quad \sum_{k=1}^{[p_q]} \gamma(k) = q .
\]

Here and below, for a finite set \( A \) a denotation \( |A| \) means the number of elements in this set. Let \( \mathcal{I}_q \) be a set of all multi-indices of order \( q \).

In addition, we enumerate all pixels \((m, n) \in D_\omega\) as to obtain a finite sequence \( \{(m_k, n_k)\}_{k=1}^{|D_\omega|} \).

Then the polynomial features of order \( q \) are defined as

\[
\Psi(\omega, \theta) = \sum_{\gamma \in \mathcal{I}_q} \sum_{p=0}^q \theta(\gamma) \prod_{k=1}^{[p_q]} (\omega(m_k, n_k))^{\gamma(k)}
\]

(1)

where \( \theta(\gamma): \mathcal{I}_q \rightarrow (-\infty; +\infty) \) are some coefficients at the corresponding terms with a multi-index \( \gamma \).

The set of such coefficients specifies a specific polynomial sign of the image. The coefficients should be matched with a specific applied problem, choosing polynomial features that best describe the image.

We showed in [9] that when imposing natural constraints on the features (1) related to the invariance to a shift and some other transformations, it turns out that the from of the whole family of polynomial features which makes sense to consider is only quadratic features in the form
\[ \Psi_2(\omega, \theta) = \sum_{(\Delta m, \Delta n) \in W_d} \theta(\Delta m, \Delta n) \hat{R}_\omega(\Delta m, \Delta n), \]  
(2)

where \( R_\omega(\Delta m, \Delta n) \) are the readings of the normalized autocovariance function of the image \( \omega(m, n) \), i.e.

\[ \hat{R}_\omega(\Delta m, \Delta n) = \frac{1}{D_\omega(\Delta m, \Delta n)} \sum_{(m, n) \in D_\omega(\Delta m, \Delta n)} (\omega(m, n) - \bar{\omega})(\omega(m + \Delta m, n + \Delta n) - \bar{\omega}) \]

Here \( W_d = [-d; +d]^2 \cap \mathbb{Z}^2 \) is a squared window of radius \( d \in (0; +\infty) \cap \mathbb{Z} \), \( D_\omega(\Delta m, \Delta n) = \{(m, n) \in D_\omega | (m + \Delta m, n + \Delta n) \in D_\omega \} \) is a set of pixels \( (m, n) \) from the region \( D_\omega \) such that the corresponding shifted on \( (\Delta m, \Delta n) \) pixels also lie in region \( D_\omega \), \( \bar{\omega} \) is a mean image gray level \( \omega(m, n) \), i.e.

\[ \bar{\omega} = \frac{1}{|D_\omega|} \sum_{(m, n) \in D_\omega} \omega(m, n) \]

The peculiarity of the quadratic features (2) is the possibility of automatic adjustment with a specific learning sample by matching the optimal parameters \( \theta(\Delta m, \Delta n) \) under which some quality function of the feature space takes the maximal value. For this, we propose to use optimization methods.

2.2. Feature space quality criteria

In the binary classification problem, the entire set of possible images \( \Omega \) is divided into two classes \( \Omega_1 \) and \( \Omega_2 \), so that \( \Omega_1 \cup \Omega_2 = \Omega \) and \( \Omega_1 \cap \Omega_2 = \emptyset \). Let \( \Phi(\omega) : \Omega \rightarrow \{\Omega_1, \Omega_2\} \) is a perfect recognition operator that maps an image to its class. To solve the recognition problem it is required to construct another operator \( \hat{\Phi}(\omega) : \Omega \rightarrow \{\Omega_1, \Omega_2\} \) which makes the decision to assign an image to one of the classes using the finite learning sample \( U \subset \Omega \) which contains the images of known class. This operator is constructed as a superposition of two operators \( \hat{\Phi}(\omega) = C(\Psi(\omega)) \), where \( \Psi(\omega) : \Omega \rightarrow \Xi \) calculates vector \( x \) of \( K \) quadratic features (2) by image \( \omega(m, n) \), and \( C(x) : \Xi \rightarrow \{\Omega_1, \Omega_2\} \) is a classifier deciding to assign a feature vector to one of the classes. Denotation \( \Xi \subseteq (-\infty; +\infty)^K \) means a feature space.

For a given set of features \( \Psi(\omega) \) it is possible to estimate the quality of the feature space by the feature vectors from the learning sample \( \{\Psi(\omega) | \omega \in U\} \). In this paper, the following quality criteria of a feature space are investigated:

1. The proportion of correctly recognized images calculated using leave-one-out cross-validation [5] by the learning sample \( U : \)

\[ \tilde{J} = \frac{1}{|U|} \left| \{\omega \in U | \Phi(\omega) = \hat{\Phi}(\omega)\} \right| \]

(3)

2. Bhattacharyya distance:

\[ \mu = \frac{1}{8} (a_1 - a_2)^2 \left( \frac{R_1 + R_2}{2} \right)^{-1} (a_1 - a_2) + \frac{1}{2} \ln \left( \frac{1}{2} \sqrt{\frac{R_1 + R_2}{2}} \right) \]

(4)
where $a_1$ and $a_2$ are the intraclass means, $R_1$ and $R_2$ are the intraclass covariance matrices:

$$a_i = \frac{1}{|U \cap \Omega|} \sum_{x \in U \cap \Omega} \Psi(x), \quad R_i = \frac{1}{|U \cap \Omega|} \sum_{x \in U \cap \Omega} \left( \Psi(x) - a_i \right) \left( \Psi(x) - a_i \right)^T.$$

3. The fourth criterion for discriminant analysis from [5]:

$$J_4 = \text{tr}^{-1}(R_2) \text{tr}(R),$$

where $R_2 = 0.5(R_1 + R_2)$ is a mean intraclass covariance matrix and $R$ is covariance matrix of the mixture of distributions.

$$R = \frac{1}{|U|} \sum_{x \in U} \left( \Psi(x) - \frac{1}{|U|} \sum_{x \in U} \Psi(x) \right) \left( \Psi(x) - \frac{1}{|U|} \sum_{x \in U} \Psi(x) \right)^T.$$

4. The criterion for discriminant analysis from [13]:

$$J_{\text{SNR}} = \frac{\|a_1 - a_2\|}{\text{tr}(R_1) + \text{tr}(R_2)}.$$

These quality criteria are used as fitness functions in procedures of optimizing the parameters of quadratic features (2).

2.3. Optimization methods for adjustment of quadratic features

We enumerate coefficients $\theta(\Delta m, \Delta n)$ and assume $\theta$ as a vector. Matching the features (2) with the learning sample $U$ is to select a vector $\theta$ under which a certain quality criterion of the feature space $J(\theta)$ returns the maximal value. To do this, you can use one of the following optimization methods.

1. Random search. Select $N$ vectors $\theta$ at random with uniform distribution in some some compact space $\Theta$. After that, one of them is chosen, for which $J(\theta)$ returns the maximal value.

2. Simulated annealing. Define the mutation operator $\zeta(\theta)$ as a Gaussian random vector, with a mathematical expectation $\theta$ and identity covariance matrix. Initially, set the zero-padded current vector $\theta$. At the each $k$-th iteration receive a new vector $\theta' = \zeta(\theta)$ and consider it a new current vector $\theta'$ if $J(\theta') > J(\theta)$. Else if $J(\theta') \leq J(\theta)$ then consider $\theta'$ a new current vector all the same with a probability $p = \exp(-J(\theta) - J(\theta'))/t_k$ where $t_k = t_{k-1}/k$ or leave $\theta$ as the current vector otherwise. Parameter $t_k$ is called an initial temperature.

3. Genetic algorithm. Define the crossover operator $c(\theta_1, \theta_2)$ as a random vector $\theta$ having $j$-th component uniformly distributed on the segment $[\min(\theta_1(j), \theta_2(j)); \max(\theta_1(j), \theta_2(j))]$. Choose an initial population of $M$ vectors at random. At each iteration, cross all pairs of vectors from the current population, after which subject each resulting vector to a mutation with probability $p_{\text{mut}}$. At the end during the transition to the next iteration, leave only $M$ vectors for which the quality criteria $J(\theta)$ returns the maximal value.

4. Hybrid algorithm. If you think about the general principles of the work of most iterative algorithms for global optimization, you can see that they are all engaged in calculating the values of the fitness function at certain points, and then simply select the point for which the calculated value of the fitness function was optimal. The essential differences between them are only in the way that they select each next for calculating the values of the fitness function. Each of the described iterative algorithms has its advantages and disadvantages. So random search algorithms have no problems with cycling at the local maximum, but they do not use information about the previously calculated values.
of the function. The genetic algorithm and the annealing simulation algorithm make a much more meaningful choice of the point, which increases the accuracy of the optimization method, but can stop at the local maximum, despite they have options to exit from it in the form of a mutation procedure in the genetic algorithm and transition to a less optimal point with high temperature of the annealing simulation algorithm.

General scheme of the iterative algorithm of global optimization for the fitness function $J(\theta)$ in the constraints $\theta \in \Theta$ with a given number of iterations $N$ looks as follows:

```python
optimize_function(J(\theta), \Theta, N)
1 A ← ∅
2 for i ← 1 : N
3 \theta ← get_next_point(A, \Theta)
4 A ← A ∪ {\theta}
5 return \arg\max_{\theta \in A} J(\theta)
```

Here we use a pseudo-code similar to that used in [14].

Varying the way to select the next point, you can get different iterative algorithms for global optimization. It is proposed at the each iteration of the hybrid algorithm to select the algorithm for selecting the next point equally among the three algorithms corresponding to the three optimization methods considered earlier. A similar idea was described in [15] for other optimization algorithms.

a) The next point is chosen randomly from the uniform distribution at $\Theta$. Previously selected points are not taken into account.

b) The next point is chosen as a mutation $\varsigma(\theta)$ of the random point $\theta \in A$. The greater the value of $J(\theta)$, the more probability to mutate.

c) The next point is chosen as a crossover $c(\theta_1, \theta_2)$ of two randomly selected points $\theta_1$ and $\theta_2$ from the set $A$ of previously selected points. The points for which the value of the function $J(\theta)$ is greater crossover more probably.

3. Experimental studies

The effectiveness of the proposed methods for matching quadratic features was studied on the following three sets of images.

1. Photos of rice from the open base of textual images Kylberg Texture Dataset [16] (320 images, resolution 576 × 576, figures 1a and 1b).

2. X-ray images of the human femoral neck obtained in clinics of the Samara State Medical University during clinical trials (95 images, resolution 1040 × 860, figures 1c and 1d).

3. Two-dimensional sections of the lungs CT, obtained in clinics of the Samara State Medical University during clinical trials (160 images, resolution 140 × 200, figures 1e and 1f).

In each series of experiments, a set of data, a quality criterion of the feature space, and an optimization method were chosen. The sample was randomly divided into a learning sample $U$ and a test sample $\bar{U}$ approximately in half, so that in both samples the proportion of images of each class was approximately the same. Next, the coordination of the quadratic features (2) with the learning sample was carried out by solving the optimization problem using the quality criterion of the feature space as the fitness function. After that, the selection of features was carried out using an algorithm, described in detail in [17]. Finally, the recognition of images from the test sample $\bar{U}$ was performed, and for this sample the proportion of correctly classified images (3) was calculated.

The nearest neighbor algorithm was used as the classification algorithm [5]. For elements of vectors, elementwise normalization was performed, so that the average value of each feature in the learning sample became zero, and the variance became unit. This procedure is described in detail in
Region of interest $D_{\omega}$ on all images except rice images were selected manually. The specific parameters of the previously described algorithms are as follows: $d = 2$ is a radius of window $w_d$ in which the features are calculated, $K = 13$ is the initial number of features, $N = 200$ is the number of optimization algorithms iterations, $M = 10$ 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 in the simulated annealing.

![Figure 1. Examples of images on which experiments were performed: white rice (a), arborio rice (b), normal femoral neck (c), femoral neck affected by osteoporosis (d), normal CT of lungs (e), CT of lungs affected by emphysema (f).](image)

Table 1 summarizes the results of a study of the quality of the first three optimization methods. For each of them and for each of the three sets of data, the best quality criterion of the feature space is shown, which provides the least probability of erroneous recognition, the least probability of erroneous recognition itself, and the least number of features $K$ remaining after the procedure of feature selection. We can see that rice images are recognized without errors, and the best results for the remaining two sets of data are achieved using different optimization algorithms and different fitness functions.
Table 1. The results of the study of methods for adjustment of the quadratic features.

| Optimization methods       | Rice    | Bones   | CT of lungs |
|----------------------------|---------|---------|-------------|
|                            | $J$     | $K$     | $J$         | $K$     | $J$     | $K$     |
| Random search              | 0.00    | 1       | 0.14        | 2       | 0.06    | 6       |
| Simulated annealing        | 0.00    | 1       | 0.06        | 4       | 0.15    | 7       |
| Genetic algorithm          | 0.00    | 1       | 0.06        | 5       | 0.12    | 4       |

Table 2. The results of the study of the hybrid optimization method.

| Criterion   | Rice    | Bones   | CT of lungs |
|-------------|---------|---------|-------------|
|             | $J$     | $K$     | $J$         | $K$     | $J$     | $K$     |
| $J$         | 1.00    | 0.00    | 0.89        | 0.22    | 0.78    | 0.20    | 2        |
| $J$         | 1.88    | 0.00    | 0.40        | 0.18    | 0.25    | 0.06    | 6        |
| $J$         | 4.34    | 0.00    | 1.43        | 0.06    | 1.32    | 0.15    | 4        |
| $J_{SNR}$   | 3.82    | 0.01    | 0.80        | 0.14    | 0.63    | 0.17    | 9        |

Table 2 details the results of a study of the efficiency of a hybrid optimization algorithm for adjustment of quadratic features. For each quality criterion of the feature space and for each data set, the greatest value of the fitness function, corresponding to the probability of erroneous recognition of images from the test sample, as well as the number of features remaining as a result of the procedure for the feature selection are shown. It can be seen that rice images are again recognized without errors and the best results for the other two sets of data are the same, but now they are both obtained using the same hybrid optimization algorithm.

The mean probability of erroneous recognition for different criteria and optimization methods is 17.23% higher than the mean error probability when using the hybrid optimization method. This means that instead of using some specific optimization method for constructing optimal quadratic features (2), we can recommend using the proposed hybrid method.

4. Conclusion

Various methods of polynomial feature adjustment for textual images were investigated. The proposed approaches are distinguished by the optimization method used and the fitness function. An original hybrid optimization algorithm was developed, combining the steps of different optimization algorithms.

In the course of experimental studies on three sets of images, the mean probability of erroneous recognition for all other optimization algorithms turned out to be 17.23% higher than the mean probability of erroneous recognition when using a hybrid algorithm for the feature adjustment. The best results obtained for all data sets using a hybrid algorithm coincide with the best results obtained with the help of other different optimization algorithms, but they were obtained using only one hybrid algorithm. Thus, we can recommend using the developed hybrid optimization algorithm to adjust the quadratic features.

The Bhattacharia distance (4) and the fourth criterion for discriminant analysis (5) show the best results from the quality criteria of the feature space. At the same time, the Bhattacharyya distance is better suited for matching features in the problem of diagnosing emphysema from CT images of the lungs, and the fourth criterion for discriminant analysis is in the problem of diagnosing osteoporosis from X-ray images of the femoral neck. Perhaps combining these functions could lead to an even more universal approach to the quadratic features adjustment.

5. References

[1] Petrou M 2006 Image processing: Dealing with texture (John Wiley & Sons, Ltd)
[2] LeCun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition Proc. IEEE 86(11) 2278-324
[3] Goodfellow I, Bengio Y and Courville A 2016 Deep learning (MIT Press)
[4] He K and Sun J 2015 Convolutional neural networks at constrained time cost The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 5353-5360

[5] Fukunaga K 1990 Introduction to statistical pattern recognition (Academic Press)

[6] Zhang G P 2000 Neural networks for classification: a survey IEEE Trans. Syst. Man Cybern. C Appl. Rev. 30(4) 451-462

[7] Goncharova E and Gaidel A 2017 Feature selection methods for remote sensing images classification CEUR Workshop Proceedings 1901 86-91

[8] Biryukova E, Paringer R and Kupriyanov A V 2016 Development of the effective set of features construction technology for texture image classes discrimination CEUR Workshop Proceedings 1638 263-269

[9] Gaidel A V 2016 Matched polynomial features for the analysis of grayscale biomedical images Computer Optics 40(2) 232-239 DOI: 10.18287/2412-6179-2016-40-2-232-239

[10] Gaidel A V 2016 Adjusted polynomial features for analysis of lung CT images CEUR Workshop Proceedings 1638 313-319

[11] Gaidel A V, Krasheninnikov V R 2016 Feature selection for diagnosing the osteoporosis by femoral neck X-ray images Computer Optics 40(6) 939-946 DOI: 10.18287/2412-6179-2016-40-6-939-946

[12] Raymond X S 1991 Elementary introduction to the theory of pseudodifferential operators (CRC Press)

[13] Yang M, Zheng H, Wang H and McClean S 2009 Feature selection and construction for the discrimination of neurodegenerative diseases based on gait analysis 3rd International Conference on Pervasive Computing Technologies for Healthcare: Pervasive Health’09 (London, United Kingdom, 1-3 April)

[14] Cormen T H, Leiserson C E and Rives R L 1990 Introduction to Algorithms (MIT Press and McGraw-Hill)

[15] Wolpert D H, Bieniawski S R and Rajnarayan D G 2013 Probability collectives in optimization Handbook of Statistics 31 61-99

[16] Kylberg G 2011 The Kylberg Texture Dataset v. 1.0: External report (Blue series) Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University 35

[17] Gaidel A V, Zelter P M, Kapishnikov A V and Khramov A G 2014 Computed tomography texture analysis capabilities in diagnosing a chronic obstructive pulmonary disease Computer Optics 38(4) 843-850

Acknowledgments
The work was partially funded by the Russian Foundation of Basic Research grant 16-41-630761 r_a and by the Russian Federation Ministry of Education and Science, and by the Federal Agency of Scientific Organizations (agreement No 007-GZ/C3363/26).