

Development of the technique for automatic highlighting ranges of interest in lungs x-ray images

N Yu Ilyasova^{1,2}, T A Chesnokova²

¹Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

²Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

e-mail: ilyasova.nata@gmail.com, c_taya@mail.ru

Abstract. In this paper, information technology has been developed for highlighting ranges of interest in lung x-ray images, based on the calculation of textural properties and classification of k-means. In some cases, the highlighted objects can describe not only the current patient's condition but also specific characteristics regarding age, gender, constitution, etc. While using the k-means method, the relationship between the segmentation error and fragmentation window size was revealed. Within the study, both a visual criterion for evaluating the quality of the segmentation result and a criterion based on calculating the clustering error on a large set of fragmented images were implemented. The study also included image pre-processing techniques. Thus, the study showed that the technology provided key objects highlighting error at 26%. However, the equalizing procedure has lessened this error to 14%. X-ray image clustering errors for fragmentation windows of 12x12, 24x24 and 36x36 were presented.

1. Introduction

Lungs X-ray is commonly performed in pulmonology. As chest x-ray is relatively fast and easy, it is particularly useful for screening respiratory diseases such as pneumonia and tuberculosis. According to medical studies, the mass of the lungs, being unevenly distributed in the chest, and has a higher density in the lower lobes. In addition, the lung mass increases with a fibrosis degree because of conjunctive tissue diffusing [1]. Many diseases can be distinguished at early stages due to the lung size screening and the ribs changes monitoring. Nowadays, there are no clinical methods to measure the density of the lungs directly. High-resolution computed tomography as the clinical standard in pulmonary fibrosis diagnostics, considers specific emission decreasing in lung tissue [2, 3]. These changes can be distinguished through the lungs x-ray image analysis.

A technique based on textural features selection has been suggested for the x-ray image analysis. Earlier, textural features had shown good results for the recognition and further diagnostics of biomedical images [4-9]. Considering the relevance of the problem, the researchers have come up with a number of solutions based on feature generation via discriminative analysis [10], followed by image segmentation [11-13]. The well-known MaZda library [14, 15] was used to calculate the textural features. The work [16] explores the detection of pathologies using blood cells images through textural characteristics of different classes of the source images in various colour subspaces; and as recent studies have shown, textural analysis is very well suited for highlighting the ranges of interest in medical pictures [4-8].

In this paper, a technology enhancing the ranges of interest, using the k-means classification methods has been proposed. The main benefit of the method is an opportunity to highlight the regions of interest by their textural features. The level of interaction between the signs, expressed by the value of correlation coefficients [10], may serve as additional information, which helps to improve the quality of the x-ray pictures of the lungs diagnosis. If the correlation between the signs is strong enough, the question arises about the appropriateness of using these signs and the possibility to reduce the dimension of feature space either by abandoning some of them and using only one of the associated signs or by identifying common properties. In the first case, we reduce the number of calculations, leaving the best feature values of any criterion. In the second case, the ability to obtain a new, more effective feature for classification appears. This technique was used in the work [17] and ensured high precision segmentation of fundus images. The presented results indicate that the proposed technology seems to be the most effective for x-ray image analysis.

The prime objective of this paper is to introduce a novel approach to identifying the lungs and background using clustering methods. In fact, as we specify it below, our approach is more than simply clustering. It is evident from the get-go that blindly using nondeterministic clustering algorithms, which typically produce (unmanageably) large numbers of local optima, would introduce great variability into the result. On the other hand, deterministic algorithms such as agglomerative hierarchical clustering typically are (substantially) slower and require essentially “guessing” the initial clustering, which in practical applications can often turn out to be suboptimal.

2. The technique for highlighting the lungs on x-ray images based on images pre-processing and k-means clustering

In this paper, we analyze the selection of the lungs at the chest x-ray images using automatic image segmentation. This method is widely used in the processing of biomedical images for various purposes, for example, for the detection of lung or liver cancer [18]. Every time we learn new about cancer, the motivation to help patients increases. Cancer differs from other diseases. It is not caused by any physical damage or biochemical imbalance. If we can identify all the major signs of the disease, this can greatly facilitate progress in understanding the origin and development of cancer. Therapeutically, if different types of cancer possess common basic structures, then the remedy for one type of cancer may be used on others. For the study, it was necessary to form a set of features that could provide the most accurate classification of x-ray images. There was a large number of textural features. In the work [19], they were divided into three groups, e.g., features based on the measurement of spatial frequencies; features based on statistical characteristics of the levels of intensity of the elements of decomposition; and features based on the description of structural elements.

Within this study, the “MaZda” program calculated 287 features. An approach based on the adjacency matrix was used [15, 19] for the formation of the features introducing a mutual disposition of pixels within the floating window. In addition, some textural properties based on the description of structural elements were considered. From the group of features based on the series lengths the following features were used, e.g. moment, inverse moment, distribution of gray levels, distribution of lengths of lines of constant optical density, the relative number of lines of constant optical density. The features mentioned above form one of the subsets out of all textural properties. Figure 1 shows the scheme for highlighting the regions of interest using the k-means method; we will describe each point in detail.

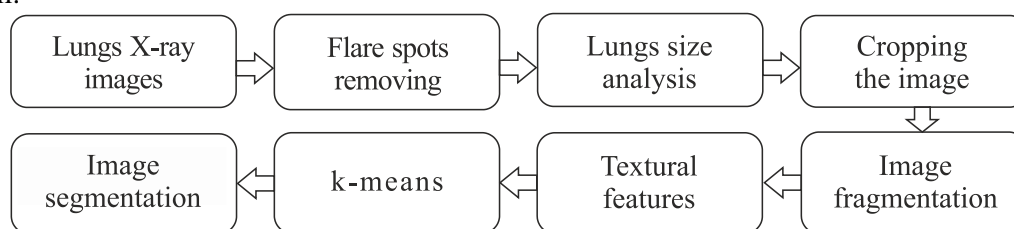


Figure 1. Technology for the regions of interest highlighting using the k-means method.

To recognize the individual objects at the x-ray picture, the following steps were undertaken. The original image was preliminary processed (Figure 2a) as follows: the flare spots with inappropriate, informative part of the image were removed; then the image was fragmented for textural features by 12×12 , 24×24 , 36×36 windows; and a set of textural features of fragmented images and clustering images based on the generated informative features was formed. For enhancing the lungs area, the initial image was preliminarily equalized. After that, the threshold processing was performed at 160, the amount of the threshold being set up experimentally. For the resulting image, see Figure 2c. An assessment of the height and width of each lung was performed, and in accordance with the data obtained (Figure 2b), nonilluminated images were cropped to leave only the parts with the lungs images being in the range of interest of the physician (Figure 2d). Next, the image was fragmented and the textural properties were calculated for every pixel by the MaZda program.

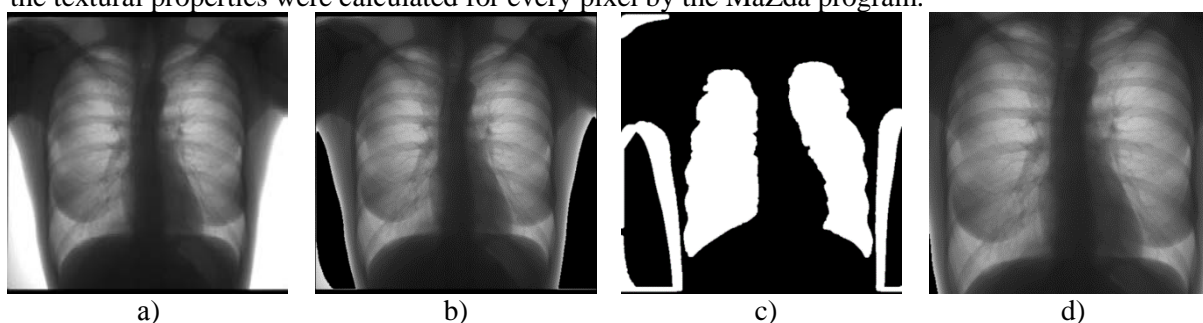


Figure 2. Image preprocessing results: (a) source image; (b) nonilluminated image; (c) equalized image with highlighted lungs and a flare spot; (d) image with the ranges of interest.

Having performed the space of features, clustering was performed based on the k-means method. The results of this procedure are illustrated in Figure 3.

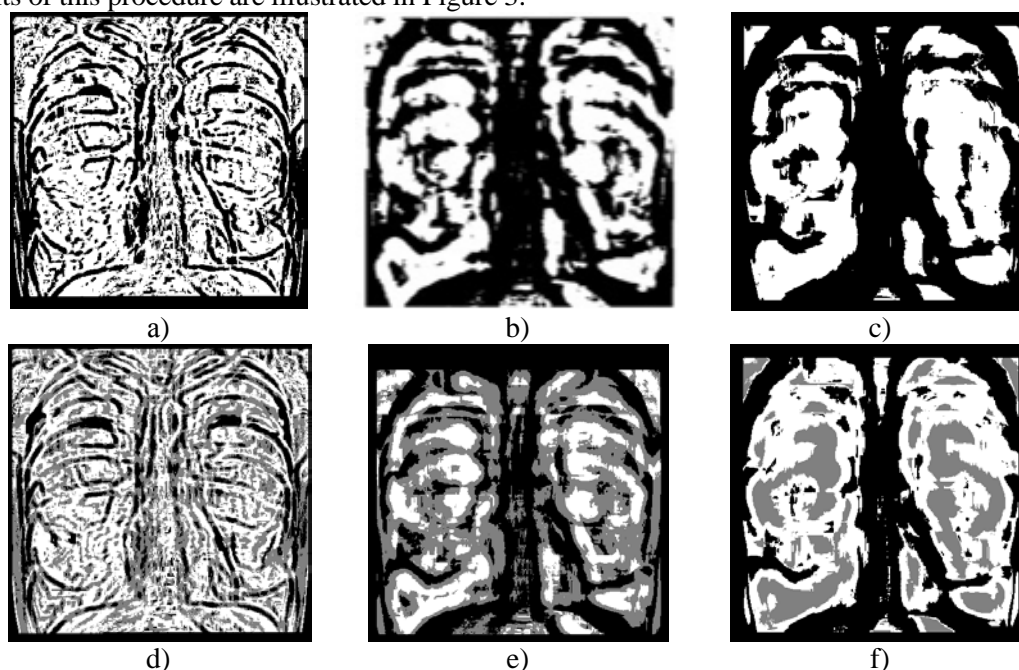


Figure 3. Results of X-ray image segmentation according to the presented technology: (a) $k = 2$, fragmentation 12×12 ; (b) $k = 2$, fragmentation 24×24 ; (c) $k = 2$, fragmentation 36×36 ; (d) $k = 3$, fragmentation 12×12 ; (e) $k = 3$, fragmentation 24×24 ; (f) $k = 3$, fragmentation 36×36 .

In the figure, the clustering of k-means on the lungs according to the textural attributes highlighted the ranges of interest incompletely. To improve the quality of the lung highlighting on X-ray pictures, the second technology based on morphological methods was considered. This technology is presented in Figure 4.

A drawback of k-means is that it is not a deterministic algorithm. Typically, there are copious local minima; and the algorithm guarantees that it will converge to a local minimum rather than to a global one. Being an iterative algorithm (as the initial centers are not pre-set in advance), k-means starts with a random set of centers at the initial iteration and converges to a different local minimum in every run. There is no magic bullet in practical applications, typically, trying to guess the location of initial centers is not easier as guessing where, for example, the global minimum is. In fact, it is acceptable in many applications. However, in the context of extracting of lungs signatures it might result in an exercise in futility. We need a way to eliminate or greatly reduce indeterminism, so we propose the second technique.

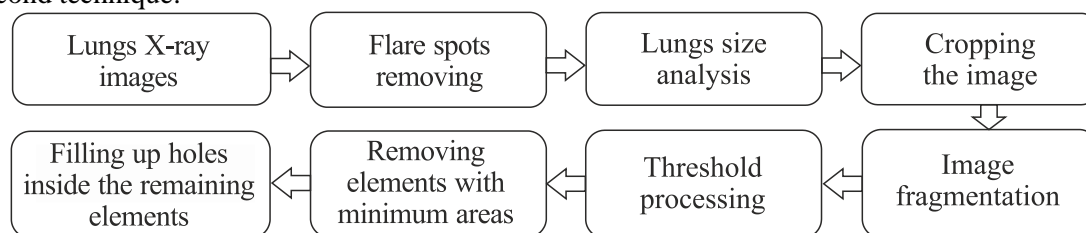


Figure 4. Technology for the selection of the ranges of interest according to the morphological methods.

At the first step, similar to the technology based on the k-means method, the original image was cropped to avoid processing fragments that were out of interest. This method contained the following steps: equalization of the original image; removing the elements with the smallest areas (while two elements, i.e. the lungs, remain); filling in holes inside the remaining elements. For the image processing result, see Figure 5.



Figure 5. Results of x-ray image processing: (a) processing by morphological operations; (b) lungs highlighted by an expert.

A comparative analysis of the results of the research methods was carried out and the corresponding errors were calculated, i.e. the k-means clustering error and the segmentation error of the equalization method. To analyze the obtained results, the image of the lungs highlighted by the expert (Figure 5b) was considered a reference. The clustering errors were calculated separately for each class, i.e. for the lungs and for the background; in case of the lungs, the error was calculated as a ratio of the number of incorrectly classified image fragments to the total number entering the region of the lungs. For the results of the analysis, see Table 1.

The total clustering error with the k-means method was 38.49% for 12×12 fragmentation, 30.61% and 28.06% for fragmentation 24×24 and 36×36 respectively. For the equalization, the error was noticeably smaller and amounted to 13.90%, and all the pixels corresponding to the segmentation error were located at the boundaries of the range of interest; that fact simplified further improvement of the algorithm.

3. Conclusion

The present study considers the application of the k-means method for the detection of images of individual objects, such as the lungs and background.

The information technology for automatic highlighting the ranges of interest on lung X-rays has been proposed. The article presents the relationship between the segmentation error and the fragmentation window size. The analysis of the obtained images was carried out, and the clustering error was found. Regarding the first method of lung X-ray, the result was less effective and revealed the objects of certain classes with an error of 28.06%, whereas technology based on the morphological methods highlighted lungs with an error of 13.90%. The clustering error of the x-ray image for the fragmentation windows of 12×12, 24×24 and 36×36 has also been found. The selection of the range of interest on the lungs x-ray in accordance with the equalization was considered preferable. The high error of the first clustering method was associated with the choice of feature space; in our case, image processing was conducted for all features calculated using MaZda software. In the future, we are planning to improve the results by the use of discriminant analysis for texture features.

Table 1. Clustering error.

	1 class (lungs)	2 class (background)	Total error
Number of correct fragments 1 nd method (12×12)	97 951	40 321	
Segmentation errors 1 nd method (12×12, %)	29.16	56.70	38.49
Number of correct fragments 1 nd method (24×24)	64 119	74 153	
Segmentation errors 1 nd method (24×24, %)	33.2	27.83	30.61
Number of correct fragments 1 nd method (36×36)	78 915	59 357	
Segmentation errors 1 nd method (36×36, %)	40.93	17.69	28.06
Number of correct fragments, 2 nd method (equalization)	73 582	72 176	
Segmentation errors 2 nd method (%)	24.26	0	13.90

4. References

- [1] Boran Z, Xiaoming Zh 2018 Lung mass density analysis using deep neural network and lung ultrasound surface wave elastography *Ultrasonics* 173-177
- [2] Mathieson J R, Mayo J R, Staples C A and Müller N L 1989 Chronic diffuse infiltrative lung disease: comparison of diagnostic accuracy of CT and chest radiography *Radiology* **1** 111-116
- [3] Verschakelen J A 2010 The role of high-resolution computed tomography in the work-up of interstitial lung disease *Curr Opin Pulm Med* **5** 503-510
- [4] Ilyasova N, Paringer R and Shirokanev A 2017 A smart feature selection technique for object localization in ocular fundus images with the aid of color subspaces *Procedia Engineering* **201** 736-745
- [5] HeiShun Yu, Tischler B, Qureshi M M, Soto J A, Anderson S, Dagainawala N, Li B and Buch K 2016 Using texture analyses of contrast enhanced CT to assess hepatic fibrosis *European Journal of Radiology* 511-517
- [6] Ilyasova N, Kupriyanov A and Paringer R 2018 Particular Use of BIG DATA in Medical Diagnostic Tasks *Pattern Recognition and Image Analysis* 114-121
- [7] Acharya U R, Ng E Y, Tan J H, Sree S V and Ng K H 2012 An integrated index for the identification of diabetic retinopathy stages using texture parameters *Journal of Medical Systems* 2011-2020
- [8] Hajek M, Dezortova M, Materka A and Lerski R 2006 Texture Analysis for Magnetic Resonance Imaging *Med4publishing* 234
- [9] Ilyasova N Yu 2013 Methods for digital analysis of human vascular system. Literature review *Computer Optics* **37(4)** 517-541
- [10] Ilyasova N, Kupriyanov A and Paringer R 2014 Formation of features for improving the quality of medical diagnosis based on discriminant analysis method *Computer Optics* **38(4)** 851-855

- [11] Ilyasova N Yu 2014 Evaluation of geometric features of the spatial structure of blood vessels *Computer Optics* **38(3)** 529-538
- [12] Shirokanev A S, Kirsh D V, Ilyasova N Yu and Kupriyanov A V 2018 Investigation of algorithms for coagulate arrangement in fundus images *Computer Optics* **42(4)** 712-721 DOI: 10.18287/2412-6179-2018-42-4-712-721
- [13] Khorin P A, Ilyasova N Yu and Paringer R A 2018 Informative feature selection based on the Zernike polynomial coefficients for various pathologies of the human eye cornea *Computer Optics* **42(1)** 159-166 DOI: 10.18287/2412-6179-2018-42-1-159-166
- [14] Strzelecki M, Szczypinski P, Materka A and Klepaczko A 2013 A software tool for automatic classification and segmentation of 2D/3D medical images *Nuclear Instruments and Methods In Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 137-140
- [15] Strzelecki M, Szczypinski P, Materka A, Klepaczko A 2009 MaZda – A software package for image texture analysis *Computer Methods and Programs in Biomedicine* 66-76
- [16] NikitaeV, Flury B 2010 The way to recognize images of cell texture *Biometrika* 33-41(in Russian)
- [17] Ilyasova N, Kupriyanov A and Paringer R 2017 Intelligent feature selection technique for segmentation of fundus images *7th International Conference on Innovative Computing Technology, INTECH* 138-143
- [18] Kakushadze Z, Willie Y 2017 *K-means and cluster models for cancer signatures *Biomolecular Detection and Quantification* 7-31
- [19] Andreev G A, Bazarskiy O V, Glauberman A S, Kolesnikov A I, Korzhik O V and Hlyavich Ya L 1984 Analysis and synthesis of random spatial textures *Foreign radio electronics* 3-33

Acknowledgments

This work was financially supported by the Russian Foundation for Basic Research under grant No. 16-41-630761, No. 17-01-00972, No. 19-29-01135 and by the Ministry of Science and Higher Education within the State assignment to the FSRC “Crystallography and Photonics” RAS No. 007-GZ/Ch3363/26 (theoretical results).