

## Разработка технологии автоматического выделения областей интереса на рентгеновских снимках легких

Т.А. Чеснокова<sup>1</sup>, Н.Ю. Ильясова<sup>1,2</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

<sup>2</sup>Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

**Аннотация.** В настоящей работе разработана информационная технология автоматического выделения областей интереса на рентгеновских снимках лёгких, основанная на вычислении текстурных признаков и классификации k-средних. По выделенным объектам в отдельных случаях можно характеризовать не только состояние лёгких пациента, но и его параметры: возраст, пол, телосложение и т.д. В процессе работы выявлена зависимость ошибки сегментации от размера окна фрагментации при использовании метода k-средних. В эксперименте использовались как визуальный критерий оценивания качества результата сегментации, так и критерий, основанный на вычислении ошибки кластеризации на большом наборе фрагментированных изображений. Исследование также включало применение методов предобработки изображений. Так исследование показало, что данная технология обеспечивает ошибку выделения ключевых объектов на уровне 26%. Однако при использовании процедуры эквализации ошибка уменьшается до 14%. Представлены ошибки кластеризации изображения рентгеновского снимка для окон фрагментации 12×12, 24×24 и 36×36.

### 1. Введение

Рентгеновские снимки легких обычно используются в пульмонологии. Поскольку обследование с использованием рентгена грудной клетки является относительно быстрой и дешевой процедурой, его можно использовать в качестве надлежащего диагностического инструмента для скрининга заболеваний органов дыхания таких, как пневмония и туберкулез. Согласно результатам медицинских исследований масса легких неравномерно распределена в органах, а именно с большей плотностью в нижних лепестках. Кроме того, масса лёгких увеличивается с увеличением степени фиброза (разрастания соединительной ткани) [1]. На основе отслеживания изменения размеров легких и параметров рёбер многие заболевания могут быть диагностированы уже на их ранних стадиях. В настоящее время нет клинических методов непосредственного измерения плотности массы легких. Высококачественная компьютерная томография является клиническим стандартом для диагностики фиброза легких с учетом специфических характеристик ослабления излучения ткани легких [2, 3]. Эти изменения можно обнаружить на основе анализа изображений рентгеновских снимков легких. Для анализа изображений рентгеновских снимков была предложена технология, основанная на выборе текстурных признаков. Текстуальные признаки и ранее показывали хорошие результаты при распознавании биомедицинских изображений и их дальнейшей диагностики [4-8]. Для

расчёта текстурных признаков была использована известная библиотека MaZda [9, 10]. В работе [11] исследуется задача выявления патологий по изображениям клеток крови на основе анализа текстурных признаков различных классов исходных изображений в разных цветовых подпространствах, и как показывают последние исследования, текстурный анализ очень хорошо подходит для выделения областей интереса на медицинских изображениях [4-8]. В этой работе предложена технология выделения областей интереса, разработанная на основе метода классификации изображений  $k$ -средних, работающая с рентгеновским снимком легких. Основным вкладом предложенного метода является возможность выделения области интереса на рентгеновских снимках по текстурным признакам. Такая методика была использована в работе [12, 15] и обеспечила высокую точность сегментации изображений глазного дна. Представленные результаты свидетельствуют о том, что предложенная технология лучше всего подходит для анализа рентгеновских снимков.

## 2. Технология выделения легких на изображениях рентгеновских снимков на основе предварительной обработки изображений и кластеризации $k$ -средних

В настоящей работе проводится анализ выделения лёгких на рентгеновских снимках грудной клетки на основе автоматической сегментации изображений. Данный способ широко используется в области обработки биомедицинских изображений в различных целях, например, выявления рака легких или печени [13]. Для работы было необходимо сформировать такой набор признаков, который обеспечит наиболее точную классификацию рентгеновского изображения. Существует большое количество текстурных признаков. В работе [14] их подразделяют на три группы: признаки, основанные на измерении пространственных частот; признаки, основанные на статистических характеристиках уровней интенсивности элементов разложения; признаки, основанные на описании структурных элементов.

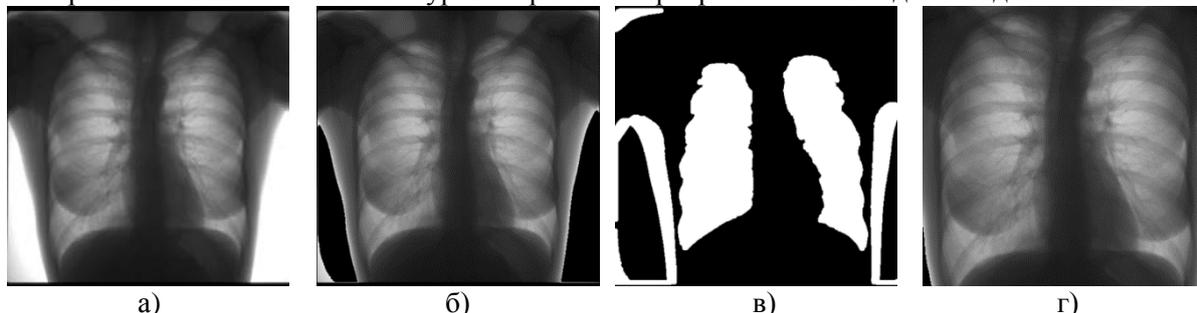
В данной работе программой «MaZda» рассчитывались 287 признаков. Для формирования признаков, учитывающих взаимное расположение пикселей внутри скользящего окна, существует подход, основанный на использовании матрицы смежности [14]. Также рассмотрены некоторые признаки текстуры, основанные на описании структурных элементов. Из группы признаков, базирующихся на длинах серий, были использованы следующие признаки: момент, обратный момент, распределение уровней серого, распределение длины линий постоянной оптической плотности, относительное число линий постоянной оптической плотности. Приведённые выше признаки являются лишь одним из подмножеств всех текстурных признаков. На рисунке 1 приведена схема выделения областей интереса методом  $k$ -средних, рассмотрим подробнее каждый из пунктов.



Рисунок 1. Технология выделения областей интереса методом  $k$ -средних.

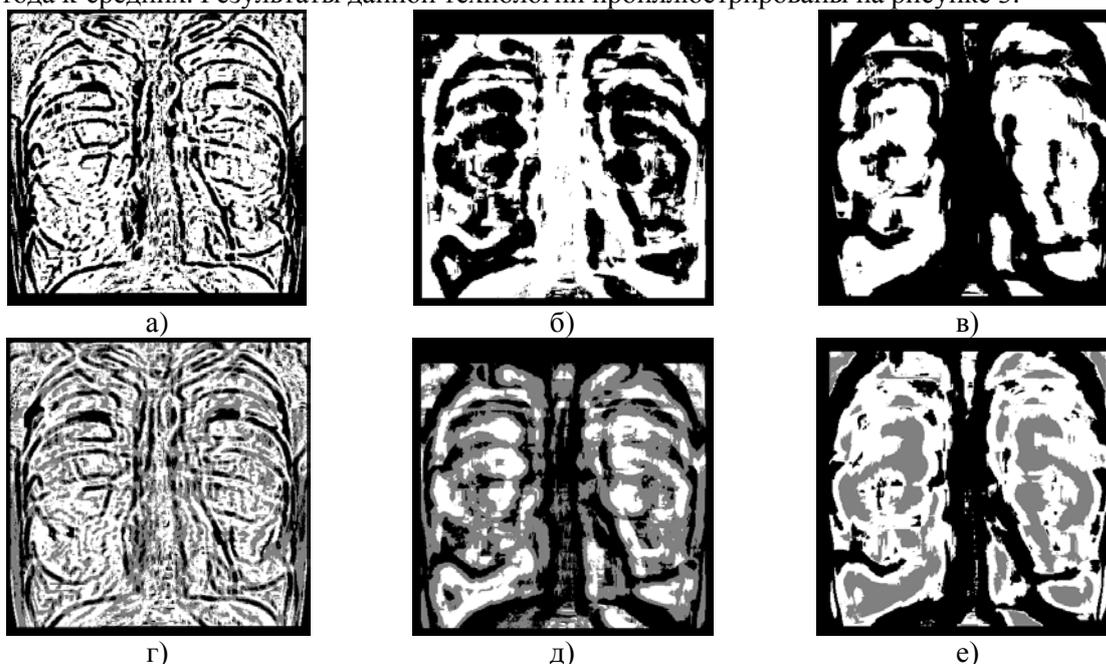
Для решения задачи распознавания отдельных объектов на изображении рентгеновского снимка были проделаны следующие шаги: предварительная обработка исходного изображения (рисунок 2а) (автоматическое удаление засветки, несоответствующей информативной части изображений); фрагментация изображения для нахождения текстурных признаков (исходное изображение фрагментировалось окнами  $12 \times 12$ ,  $24 \times 24$ ,  $36 \times 36$ ); формирование набора текстурных признаков фрагментированных изображений и кластеризация изображений на базе сформированных информативных признаков. Для выделения области с лёгкими предварительно осуществлялась эквализация исходного изображения, далее пороговая обработка с порогом 160, подобранным экспериментальным образом. Полученное изображение проиллюстрировано на рисунке 2в. После чего была

осуществлена оценка высоты и ширины каждого из лёгких, и в соответствии с полученными данными обрезано изображение без засветки так (рисунок 2б), чтобы оставить только часть с лёгкими, являющимися областью интереса врача (рисунок 2г). Далее было фрагментировано изображение и вычислены текстурные признаки программой Mazda для каждого из пикселей.



**Рисунок 2.** Результаты предобработки изображения: исходное изображение (а); изображение без засветки (б); эквализированное изображение с выделенными легкими и засветкой (в); изображение содержащее области интереса (г).

После формирования признакового пространства была произведена кластеризация на основе метода  $k$ -средних. Результаты данной технологии проиллюстрированы на рисунке 3.

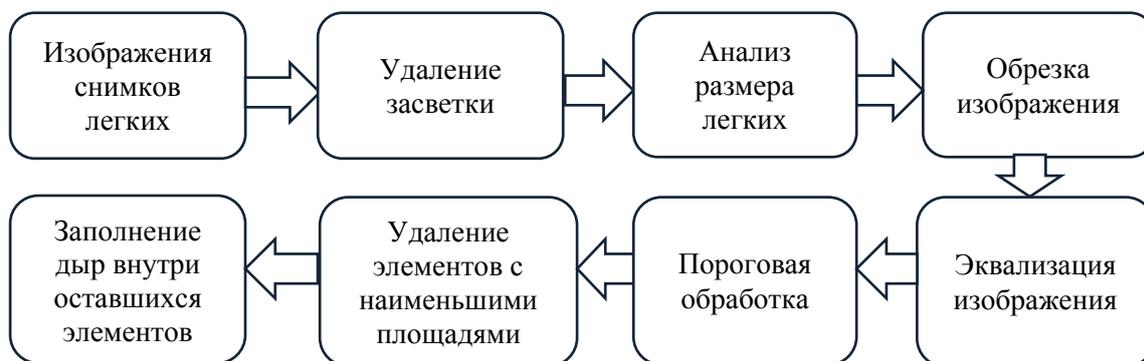


**Рисунок 3.** Результаты сегментации изображений рентгеновского снимка с использованием представленной технологии: при  $k=2$  и фрагментация  $12 \times 12$  (а); при  $k=2$  и фрагментация  $24 \times 24$  (б); при  $k=2$  и фрагментация  $36 \times 36$  (в); при  $k=3$  и фрагментация  $12 \times 12$  (г); при  $k=3$  и фрагментация  $24 \times 24$  (д); при  $k=3$  и фрагментация  $36 \times 36$  (е).

Как видно из рисунка, кластеризация  $k$ -средних на лёгких по текстурным признакам выделила области интереса не полностью. Для улучшения качества выделения лёгких на рентгеновских снимках была рассмотрена вторая технология выделения областей интереса, основанная на морфологических методах. Данная технология представлена на рисунке 4.

Как и для технологии, основанной на методе  $k$ -средних, сначала исходное изображение было обрезано таким образом, чтобы при выполнении операций не обрабатывать фрагменты, не несущие интерес. Данный метод реализуется на основе выполнения следующих шагов: эквализация исходного изображения; удаление элементов с наименьшими площадями

(остаются только 2 элемента, являющиеся лёгкими); заполнение дыр внутри оставшихся элементов. Полученный результат обработки изображения представлен на рисунке 5.



**Рисунок 4.** Технология выделения областей интереса морфологическими методами.



**Рисунок 5.** Результаты обработки изображений рентгеновского снимка: обработка с использованием морфологических операций (а); легкие, выделенные экспертом (б).

В работе был проведён сравнительный анализ результатов исследований представленных методов и вычислены соответствующие ошибки: ошибка кластеризации методом k-средних и ошибка сегментации метода, основанного на эквализации.

Для анализа полученных результатов эталонным изображением считалось изображение лёгких, выделенных экспертом (рисунок 5б). Ошибка кластеризации рассчитывалась для каждого класса отдельно, то есть для легких своя ошибка – отношение количества неверно классифицированных фрагментов изображения к общему количеству, входящему в область легких, для фона аналогично. Результаты анализа представлены в таблице 1.

**Таблица 1.** Ошибка кластеризации.

	1 класс (легкие)	2 класс (фон)
Количество верных фрагментов (12×12)	36 564	60 516
Ошибка сегментации (12×12, %)	62,36	16,15
Количество верных фрагментов (24×24)	47 812	59 833
Ошибка сегментации (24×24, %)	50,73	17,1
Количество верных фрагментов (36×36)	63 750	61 833
Ошибка сегментации (36×36, %)	34,38	14,33
Количество верных фрагментов 2 метода (эквализации)	73 582	72 176
Ошибка кластеризации (%)	24,26	0

Общая ошибка кластеризации равна 42,67% для фрагментации 12×12, 36,42% и 25,83% для фрагментации 24×24 и 36×36 соответственно при кластеризации методом k-средних. Для эквализации эта ошибка заметно меньше и составляет 13,9% и все пиксели приходящиеся на ошибку сегментации находятся на границах области интереса, что заметно упрощает дальнейшее улучшение алгоритма.

### 3. Заключение

В данной работе проведено исследование приложения метода k-средних для нахождения на снимках отдельных объектов таких, как легкие и фон.

Предложена информационная технология автоматического выделения областей интереса на рентгеновских снимках легких. В статье представлена зависимость ошибки сегментации от размера окна фрагментации. Проведен анализ полученных изображений, и найдена погрешность кластеризации. Сделан вывод о том, что работа рассматриваемого первого метода на рентгеновских снимках лёгких менее эффективна и выявляет объекты определённых классов с ошибкой 25,83% тогда, как технология, основанная на морфологических методах, выделяет лёгкие с ошибкой 13,9%. Сделан вывод о преимуществе выделения области интереса на рентгеновских снимках лёгких с помощью операции эквализации.

Также была найдена ошибка кластеризации изображения рентгеновского снимка для окон фрагментации 12×12, 24×24 и 36×36.

### 4. Литература

- [1] Zhou, B. Lung mass density analysis using deep neural network and lung ultrasound surface wave elastography / B. Zhou, X. Zhang // *Ultrasonics*. – 2018. – Vol. 89. – P. 173-177.
- [2] Mathieson, J.R. Chronic diffuse infiltrative lung disease: comparison of diagnostic accuracy of CT and chest radiography / J.R. Mathieson, J.R. Mayo, C.A. Staples, N.L. Müller // *Radiology*. – 1989. – Vol. 171(1). – P. 111-116.
- [3] Verschakelen, J.A. The role of high-resolution computed tomography in the work-up of interstitial lung disease // *Curr Opin Pulm Med*. – 2010. – Vol.16 (5). – P. 503-510.
- [4] Ilyasova, N. A smart feature selection technique for object localization in ocular fundus images with the aid of color subspaces/ N. Ilyasova, R. Paringer, A. Shirokanev // *Procedia Engineering*. – 2017. – Vol. 201. – P. 736-745.
- [5] HeiShun, Yu. Using texture analyses of contrast enhanced CT to assess hepatic fibrosis/ Yu. HeiShun, B. Tischler, M.M. Qureshi, J.A. Soto, S. Anderson, N. Dagainawala, B. Li, K. Buch // *European Journal of Radiology*. – 2016. – Vol. 85 – P. 511-517.
- [6] Ilyasova, N. Particular Use of BIG DATA in Medical Diagnostic Tasks / N. Ilyasova, A. Kupriyanov, R. Paringer // *Pattern Recognition and Image Analysis*. – 2018. – Vol. 28(1). – P. 114-121.
- [7] Acharya, U.R. An integrated index for the identification of diabetic retinopathy stages using texture parameters / U.R Acharya, E.Y. Ng, J.H. Tan, S.V. Sree, K.H. Ng // *Journal of Medical Systems*. – 2012. – Vol. 36(3). – P. 2011-2020.
- [8] Hajek, M. Texture Analysis for Magnetic Resonance Imaging / M. Hajek, M. Dezortova, A. Materka, R. Lerski // *Med4publishing*. – 2006. – P. 234.
- [9] Strzelecki, M. A software tool for automatic classification and segmentation of 2D/3D medical images / M. Strzelecki, P. Szczypinski, A. Materka, A. Klepaczko // *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. – 2013. – Vol. 702. – P. 137-140.
- [10] Szczypiński, M. MaZda – A software package for image texture analysis / M. Strzelecki, P. Szczypinski, A. Materka, A. Klepaczko // *Computer Methods and Programs in Biomedicine*. – 2009. – Vol. 94(1). – P. 66-76.
- [11] Nikitaev, V. The way to recognize images of cell texture / V. Nikitaev, B. Flury // *Biometrika*. – 2010. – Vol. 77. – P. 33-41.
- [12] Ilyasova, N. Intelligent feature selection technique for segmentation of fundus images / N. Ilyasova, A. Kupriyanov, R. Paringer // *7th International Conference on Innovative Computing Technology, INTECH, 2017*. – P. 138-143.
- [13] Kakushadze, Z. \*K-means and cluster models for cancer signatures / Z. Kakushadze, W. Yuc // *Biomolecular Detection and Quantification*. – 2017. – Vol. 31. – P. 7-31.
- [14] Андреев, Г.А. Анализ и синтез случайных пространственных текстур / Г.А. Андреев, О.В. Базарский, А.С. Глауберман, А.И. Колесников, Ю.В. Коржик, Я.Л. Хлявич // *Зарубежная радиоэлектроника*. – 1984. – С. 3-33.

- [15] Широканев, А.С. Исследование алгоритмов расстановки коагулятов на изображение глазного дна / А.С. Широканев, Д.В. Кириш, Н.Ю. Ильясова, А.В. Куприянов // Компьютерная оптика. – 2018. – Т. 42, № 4. – С. 712-721. DOI: 10.18287/2412-6179-2018-42-4-712-721.

### **Благодарности**

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (гранты № 16-41-630761, № 17-01-00972, № 18-37-00418), государственного задания 3.3025.2017/4.6 и Министерства науки и высшего образования Российской Федерации, в рамках выполнения работ по Государственному заданию ФНИЦ «Кристаллография и фотоника» РАН (соглашение №007-ГЗ/ЧЗ363/26).

## **Development of automatic selection technique of interest regions in lungs x-rays images**

**Т.А. Chesnokova<sup>1</sup>, N.Yu. Piyasova<sup>1,2</sup>**

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

<sup>2</sup>Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

**Abstract.** We propose a technique for automatic selection technique of interest regions in lungs x-rays images. The relevance of the problem is associated with enhancing the lung disease count. The technique is based on the texture analysis of lungs x-rays images. The automatic selection technique of interest regions is performed using the k-means clustering algorithm and the morphological operations. The best values of image fragmentation dimensions for the image segmentation required for regions of interest are determined herein. The experiment used both a visual criterion for evaluating the quality of the segmentation result, and a criterion based on the calculation of the clustering error on a large set of fragmented images. The study also included the use of image preprocessing techniques. So the study showed that this technology provides an error in the selection of key objects at a level of 26%. However, when using the equalization procedure, the error is reduced to 14%. X-ray image clustering errors for fragmentation windows 12×12, 24×24 and 36×36 are presented.