

**АЛГОРИТМ ОБНАРУЖЕНИЯ ОБРАЗОВАНИЙ НА СНИМКАХ КОМПЬЮТЕРНОГО  
ТОМОГРАФА С ИСПОЛЬЗОВАНИЕМ КЛАСТЕРИЗАЦИИ**

Максимова Е.И.

Научный руководитель: Хаустов П.А., ассистент  
Национальный исследовательский Томский политехнический университет,  
Россия, г.Томск, пр. Ленина, 30, 634050  
E-mail: YelenaMaksimova@yandex.ru

**ALGORITHM FOR LUNGS FORMATIONS DETECTION IN X-RAY COMPUTER TOMOGRAPHY  
SCANS BASED ON CLUSTERING**

Maksimova Y.I.

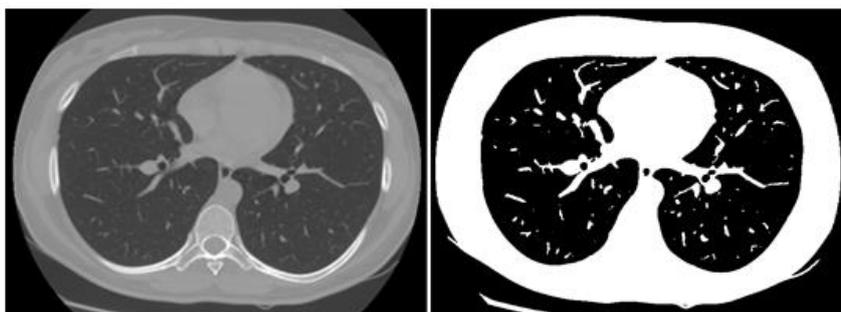
Scientific Supervisor: Khaustov P.A., assistant  
Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050  
E-mail: YelenaMaksimova@yandex.ru

The image processing algorithm of searching formation in the lungs has been proposed for computer tomography scans. The approach is based on clustering to detect some regions in the image. Different heuristics were suggested to detect lungs on the image and to discard some objects that are not to be detected by the algorithm. The algorithm has been implemented and tested for the real image database of x-ray computer tomography. Future improvements for the proposed approach were suggested.

**Введение.** Компьютерная томография – это метод послойной диагностики организма, основанный на рентгеновском излучении. Современные компьютерные томографы (КТ) позволяют получать изображения с высоким пространственным разрешением за короткий промежуток времени и в медицинских целях используются для выявления или уточнения патологических процессов в организме. В отличие от обычного рентгена, на котором лучше всего видны кости, при помощи КТ можно исследовать так же и мягкие ткани. Однако затраты на оборудование и программное обеспечение (ПО) для томографии велики по сравнению с обычным рентгеновским аппаратом. Государственного финансирования недостаточно, чтобы обеспечить все медицинские учреждения страны достаточным количеством КТ, ввиду чего пациенты вынуждены ждать проведения обследования, что зачастую негативно сказывается на процессе выздоровления. При этом в комплект ПО для КТ, приобретаемых российскими медицинскими учреждениями, не входят программы, предназначенные для анализа снимков (обнаружения образований) КТ.

Ввиду значительно увеличения числа больных на территории РФ, нуждающихся в проведении различного рода обследований при помощи КТ, актуальной задачей является повышение скорости проведения подобных обследований путем частичной автоматизации процесса анализа снимков [1].

**Предложенный метод.** Одной из удобных особенностей изображений компьютерного томографа является возможность их представления в градациях серого, что существенно упрощает последующую обработку изображения. Для выполнения дальнейших стадий обнаружения легочных долей и образований в них выполняется операция адаптивной бинаризации изображения методом Оцу. Пример результата выполнения такой бинаризации представлен на рис. 1:



*Рис. 1. Изображение здорового легкого и результат его бинаризации*

После того, как изображение было бинаризовано, изображение можно разделить на различные кластеры черного и белого цветов. Среди этих кластеров необходимо выделить легочные доли. В основу метода поиска легочных долей были заложены некоторые эвристические утверждения: искомые доли представлены кластерами черного цвета, должны быть расположены примерно на одной высоте, симметрично относительно центра, имея схожие размеры. Другими словами, можно рассмотреть все пары черных кластеров и, используя предложенные эвристики, определить пару кластеров, которая описывает две легочные доли. Для этого потребуются решить задачу многокритериальной оптимизации для численных характеристик, соответствующих всем предложенным эвристикам. Для этого для каждой пары кластеров черного цвета необходимо определить три вещественных значения:

- Модуль разности между ординатами центров масс кластеров пары (разность высот);
- Модуль разности размеров кластеров пары (разность размеров);
- Модуль разности расстояний от центра масс до середины изображения (величина характеризующая симметрию).

После чего, можно упорядочить все пары по каждому из этих значений в отдельности в порядке уменьшения числового значения. После чего можно просуммировать для каждой из пар порядковые номера этой пары во всех полученных списках. Чем меньше полученная сумма, тем больше данная пара соответствует описанию легочных долей. В качестве дополнительного критерия при равенстве полученных сумм можно использовать суммарный размер пары кластеров – легочные доли занимают существенную площадь снимка компьютерного томографа, следовательно, два наиболее крупных кластера, имеющих наименьшее значение посчитанной суммы порядковых номеров, являются легочными долями.

Так как при определении цвета кластера используется средний цвет всех пикселей, принадлежащих кластеру, пиксели на границах этого кластера могут быть намного светлее усредненного значения. Группы таких пикселей могут быть ложно распознаны как образования в легких. Для того, чтобы избежать подобного эффекта необходимо удалять внешние слои пикселей кластеров до тех пор, пока менее 75% пикселей внешнего слоя имеют яркость намного меньше средней по кластеру.

Внутри черного кластера, соответствующего легочной доли, необходимо обнаружить все белые кластеры, среди которых могут быть как образования в легких, так и элементы легочного рисунка. Тип каждого из таких кластеров необходимо определить: элемент легочного рисунка, образование в легких, легочная жидкость и другие.

В виду графической особенности легочного рисунка можно заметить, что площадь охватывающего прямоугольника элемента легочного рисунка намного больше, чем количество пикселей в этом элементе. Если же это утверждение неверно для такого элемента, то соотношение горизонтального и вертикального размеров охватывающего прямоугольника далеко от единицы (один из этих размеров существенно больше другого). Это объясняется тем, что элементы легочного рисунка, как правило, являются продолговатыми группами светлых пикселей, у некоторых из которых имеются схожие продолговатые ответвления. С использованием эмпирических формул можно выявить закономерность в соотношении площади охватывающего прямоугольника и площади самого элемента легочного рисунка: чем больше пикселей содержит кластер, тем меньше должно быть соотношение большего размера к меньшему. Заполненностью кластера будем называть процент пикселей охватывающего прямоугольника этого кластера, принадлежащих непосредственно этому кластеру. В зависимости от количества пикселей в кластере, эмпирическим путем можно определить необходимую пороговую величину заполненности для корректного распознавания элементов легочного рисунка. График зависимости порогового значения заполненности от количества пикселей в кластере представлен на рис. 2:

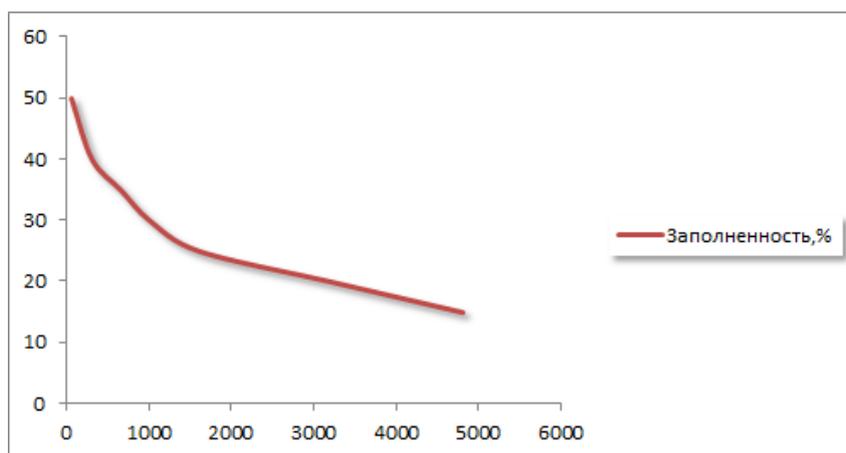
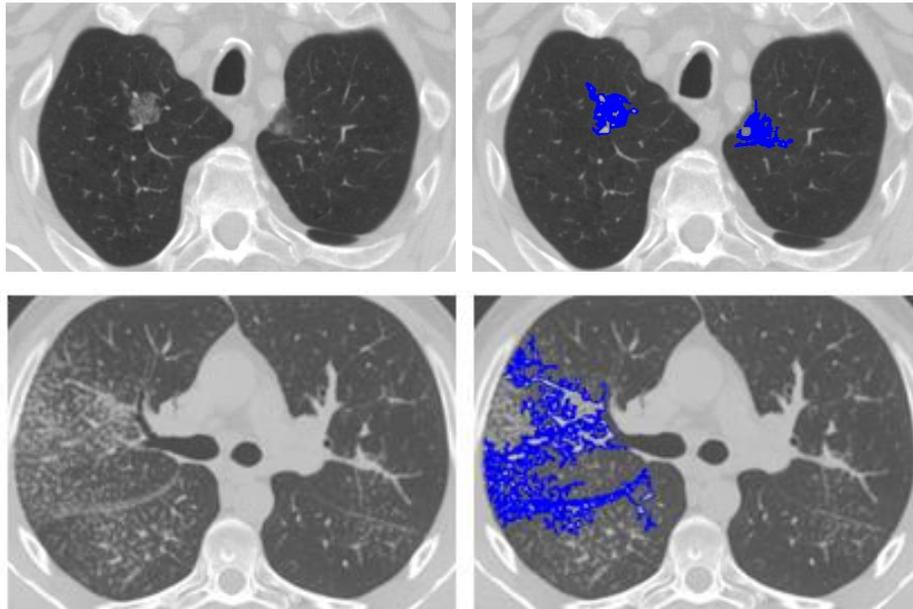


Рис. 2. Графическая зависимость заполненности от размера кластера

Для улучшения качества классификации также возможно использовать и сторонние классификаторы: средства нечеткой логики, деревья решений, искусственные нейронные сети и сети, основанные на использовании радиально-базисных функций [2].

При реализации алгоритма разбиения бинаризованного изображения на кластеры использовался волновой алгоритм (алгоритм обхода в ширину для компонент четырех-связности). [3] Результатом работы алгоритма является изображение аналогичное исходному, на котором синим цветом обведены объекты, распознанные как образования в легких. Примеры результатов работы алгоритма представлены на рис. 3:



*Рис. 3. Примеры результатов обнаружения образований в легких*

Как можно заметить, на приведенных выше снимках компьютерного томографа изображены два случая образований в легких. В первом случае образование имеет незначительный размер, во втором же – занимает почти целую легочную долю. В обоих случаях алгоритмом правильно обнаружены и распознаны образования в легочных долях, а элементы легочного рисунка правильно классифицированы и не выделены синим цветом.

Вычислительная сложность алгоритма линейно зависит от количества пикселей на изображении, что говорит о возможности применения такого алгоритма в системах, обладающих высоким быстродействием.

**Полученные результаты.** Для апробации предложенного алгоритма была использована база изображений одного из медицинских учреждений томской области. На выборке из двадцати изображений алгоритм безошибочно определяет отсутствие образований или же обнаруживает сами образования или их большую часть на снимке.

В будущем планируется расширить используемую базу изображений компьютерного томографа новыми образцами. В том числе планируется добавление изображений с другими видами образований в легких. Для улучшения качества распознавания планируется разработать классификатор на основе искусственной нейронной сети с целью распознавания легочного рисунка, легочных образований и других объектов на изображениях компьютерного томографа.

#### СПИСОК ЛИТЕРАТУРЫ

1. Андреева О.В., Исакова Л.Е. Инструменты повышения эффективности деятельности медицинских учреждений // Проблемы управления здравоохранением. – 2002. – № 1. – С. 79-82
2. Böhm, S., Kailing, K., Kröger, P., Zimek, A. Computing Clusters of Correlation Connected objects // Proceedings of the 2004 ACM SIGMOD international conference on Management of data. – 2004. – № 1. – p. 132-136
3. Вятчин Д.А. Нечёткие методы автоматической классификации. – М.: Технопринт, 2004. – 219 с.