

УДК 004.9

В.А. Ковалев, В.А. Левчук

**КАРТИРОВАНИЕ ХАРАКТЕРИСТИК СВЕРХБОЛЬШИХ
ГИСТОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЙ РАКОВОЙ ТКАНИ**

Предлагается метод цифровой обработки гистологических изображений сверхбольшого размера порядка $100\,000 \times 100\,000$ пикселей. Приводится способ построения карт распределения различных параметров по изучаемому образцу ткани. Получаемые карты пригодны как для количественного, так и для визуального анализа специалистами-патологами. Проводится корреляционный анализ между вычисленными параметрами изображений опухолевых тканей и клиническими данными по пациентам.

Введение

Известно, что визуальное исследование гистологических изображений, снятых с образцов ткани, является важным инструментом в окончательной диагностике, определении стадии и оценке результатов лечения различных типов рака [1, 2]. Между тем область анализа цифровых гистологических изображений на данный момент еще недостаточно изучена и менее развита по сравнению с другими отраслями современной обработки изображений [3–5]. Одной из причин такого положения является высокая морфологическая сложность гистологических изображений [3, 4, 6]. Проблема усугубляется еще больше с появлением новых методов предварительной гистохимической обработки («окрашивания») образцов ткани, а также с переходом к съемке изображений всего образца ткани под большим увеличением. Последний фактор приводит к получению сверхбольших изображений с размерами сторон до $100\,000$ пикселей и объемом, достигающим величины порядка $10\,000$ мегапикселей [7].

На сегодняшний день стандартным подходом к извлечению информации о процессах лимфо- и ангиогенеза из гистологических образцов является выделение на изображениях ткани так называемых «горячих пятен» – точек максимальной васкуляризации и последующего анализа этих сравнительно небольших изображений [8, 9]. В соответствии с данной методикой обрабатывается лишь малая часть информации, представленной на образцах раковой ткани. В настоящей работе предлагается способ извлечения полезных данных из полных изображений всего образца ткани с возможностью последующих как численной, так и визуальной обработок получаемых данных. В основе предлагаемого подхода лежит идея картирования больших изображений, т. е. разбиения их на малые части, вычисления характеристик по каждой из таких малых частей с последующим построением карты распределения вычисленных характеристик по всему анализируемому образцу ткани. При этом одному пикселу получаемой выходной карты соответствует некоторый «элементарный» участок исходного большого изображения (в нашем случае это квадрат размером $S_0 = 512 \times 512$ пикселей), по которому вычисляется значение картируемого параметра. Таким образом, размер выходной карты равен размеру входного большого изображения в пикселах, деленному на размер элементарного квадратного участка анализа S_0 .

В настоящий момент существует множество работ, посвященных разработке методик визуализации и обработки больших гистологических изображений. Многие из предлагаемых методик применяют кэширование, построение масштабных моделей, разбиение большого гистологического изображения на малые части с последующей отдельной их обработкой [10–13]. При этом результаты обработки частей изображения используются лишь для оценки численных характеристик морфологической структуры ткани без возможности визуального анализа распределения вычисленных характеристик по целому образцу. Однако человеку не под силу проанализировать всю структуру образца опухолевой ткани, просмотрев десятки тысяч малых изображений. Целью же настоящей работы является построение такой масштабной модели сверхбольшого изображения, которая отражала бы распределение интересующих нас характеристик и в то же время была пригодна для визуального анализа.

Как правило, при проведении исследований с целью нахождения связей между морфологическими структурами изображений и состоянием пациентов в распоряжении исследователя име-

ется некоторая база данных (БД) по пациентам. Она содержит как биомедицинские изображения разных модальностей, так и различные характеристики пациентов, например общие социальные данные, клинические наблюдения, результаты лабораторных анализов, данные по предыдущим заболеваниям самого пациента, его родителей и близких родственников. Соответственно окончательная задача заключается в том, чтобы выделить количественные признаки, описывающие пространственную структуру изображений и провести корреляционный анализ этих признаков и данных о пациентах из БД с целью выявления статистически значимых связей.

1. Особенности гистологических изображений

Исходная БД содержала записи данных и гистологические изображения тканей по 68 пациентам, страдающим раком яичников (женщины, средний возраст 60 лет, $STD = 11$). По каждому пациенту имелись гистологические снимки образцов ткани, обработанной иммуногистохимическим препаратом D2-40 для маркировки лимфатических сосудов. Изображения были получены с помощью современной автоматической сканирующей системы Leica SCN400 при увеличении $\times 400$ и имели размеры порядка $100\,000 \times 100\,000$ пикселей. Для картирования признаков указанные большие изображения разрезались на квадратные блоки размером 512×512 пикселей, по каждому из которых вычислялось числовое значение рассматриваемого параметра (рис. 1). После этого строились карты распределения значений изучаемого параметра по образцу (срезу) исследуемой ткани. Таким образом, каждому пикселу выходной карты соответствовал участок изображения размером 512×512 пикселей, а его цвет определялся соответствующим значением изучаемого параметра. Такие карты дают возможность визуально оценить исследуемые морфологические параметры всего большого изображения образца ткани одновременно, а также использовать их количественные значения для последующего компьютеризированного анализа статистических связей с теми или иными данными, характеризующими состояние пациента.

Данные о состоянии пациентов включали в себя около 80 различных характеристик, таких, как международная классификация степени прогресса раковой опухоли TNM, отметки о предыдущих перенесенных болезнях, которые могут иметь отношение к исследуемому онкологическому заболеванию, степень распространения опухолевого процесса, хирургические данные, данные о проведенных сеансах химиотерапии, текущий статус жив/умер для каждого пациента и т. д.

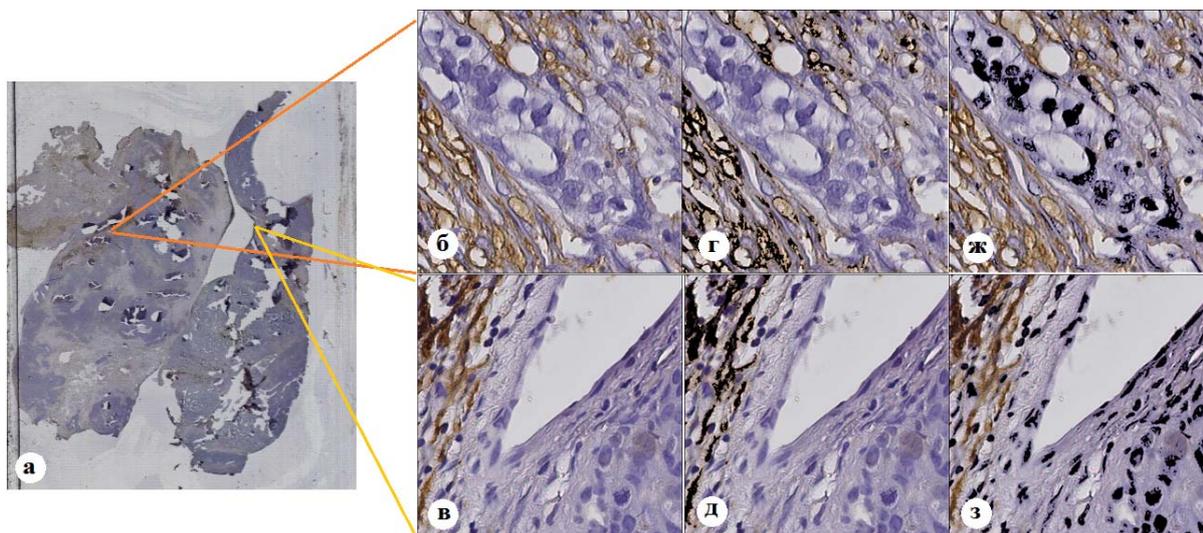


Рис. 1. Примеры гистологических изображений: *а)* целого образца ткани; *б), в)* отдельных его блоков на рабочем увеличении; *г), д)* тех же участков с выделенными автоматически сосудами и *ж), з)* с выделенными ядрами клеток

2. Алгоритм картирования сверхбольших изображений

Определим требования, предъявляемые к методу извлечения количественных характеристик из сверхбольших гистологических изображений.

В настоящей работе используется БД по пациентам, содержащая десятки тысяч малых изображений размером 512×512 по каждому пациенту, а общее количество изображений, подлежащих обработке, близко к миллиону. В случае столь объемной базы изображений теоретически весьма затруднительно, а в реальных клинических условиях и практически невозможно использовать сложные методы, требующие значительного времени для вычисления необходимых параметров. Таким образом, первое требование, предъявляемое к разрабатываемой методике анализа, – это ее простота и высокая скорость работы.

Вместе с тем следует отметить, что ввиду огромного количества изображений, подлежащих обработке, достижение высокой точности сегментации каждого из них не является приоритетной целью исследования. Поскольку один малый участок изображения соответствует всего одному пикселу результирующей карты, ценными и информативными являются данные, вычисленные не по каждому в отдельности, а по некой совокупности таких участков. Поставленная задача является скорее не точной, а статистической, и рассматриваемые параметры не обязательно должны быть вычислены точно, но должны быть статистически устойчивыми и существенно коррелировать с характеристиками реальных объектов, представленных на изображении.

Отсутствие требования высокой точности также позволяет не учитывать ошибки сегментации объектов, попавших на границу раздела малых блоков 512×512 , и обрабатывать каждый блок независимо от остальных.

Количество параметров, описывающих содержимое изображения, должно быть невелико, чтобы удовлетворять хорошо известному статистическому требованию – оно должно быть значительно меньше числа пациентов во избежание обнаружения ложных, случайных корреляций, т. е. псевдозависимостей.

Исходя из изложенных требований, для сегментации объектов в данной работе предлагается использовать простой метод порогового ограничения (цветовой бинаризации), дополнив его подходящей процедурой статистической корректировки результатов на основе уравнения регрессии, полученного на обучающей выборке.

В качестве характеристических объектов в данной работе использовались сети лимфатических сосудов (см. коричневые участки на рис. 1, б, в) и ядра клеток (темно-синие пятна на рис. 1, б, в). Объекты выделялись при помощи классического порогового метода по компонентам цветовой модели HSV. Для корректного определения порогов ряд изображений предварительно сегментировался вручную. Далее, переводя цвета выделенных пикселей из исходного цветового представления RGB в цветовую модель HSV (Hue, Saturation, Value), можно вначале построить гистограммы распределения пикселей по параметрам *hue*, *saturation*, *value*, а затем по гистограммам определить пороги для соответствующих значений.

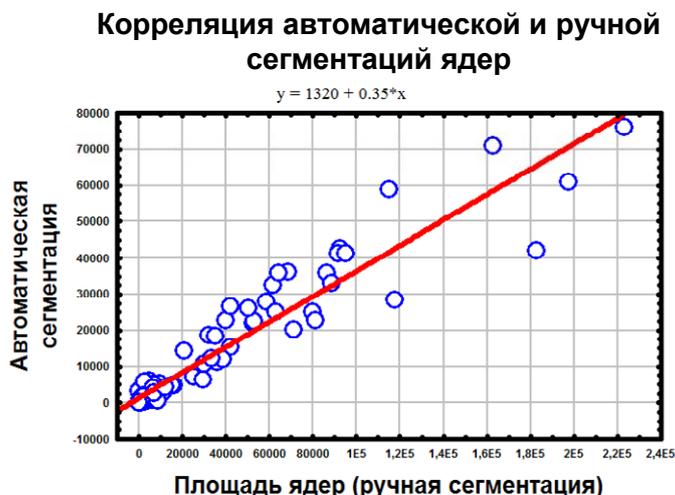


Рис. 2. Корреляция между количеством пикселей, выделенных на изображениях при помощи ручной и автоматической сегментаций

Следует убедиться, что качество автоматической сегментации при выбранных порогах является удовлетворительным и результат пороговой сегментации (после соответствующей статистической корректировки) близок к результатам ручной. Для этого необходимо провести корреляционный анализ площади выделенных при обоих способах сегментации объектов (количество пикселей), варьируя пороги, добиться максимально возможной корреляции, построить уравнение линейной регрессии и скорректировать полученный результат. На рис. 2 в качестве примера представлена регрессия результатов ручного и автоматического выделения ядер клеток, полученная на обучающей выборке из 91 изображения. Учитывая, что значение коэффициента корреляции для представленного уравнения регрессии достигает значения $R = 0,950$, качество данной методики быстрой автоматической сегментации можно признать удовлетворительным.

Таким образом, в нашем методе картирования сверхбольших изображений можно выделить следующие этапы:

– предварительную калибровку на обучающей выборке (выбор порогов, регрессионный анализ);

– разбиение сверхбольшого изображения на блоки размером 512×512 пикселей;

– вычисление по каждому блоку двух параметров: относительной площади изображения, занимаемой лимфатическими сосудами (*vess*), и относительной площади изображения, занимаемой ядрами клеток (*nuclei*);

– построение цветокодированной карты распределения характеристик сверхбольшого изображения (цвет каждого пиксела карты соответствует значению вычисленного параметра).

Построенные таким образом карты дают возможность выявлять соответствия между морфологическим строением раковой ткани и имеющимися данными о пациенте (рис. 4). Отметим, что выбор интересующих параметров и метод их извлечения из малых изображений, вообще говоря, могут быть произвольными.

Кроме визуального представления о структуре образца ткани, хотелось бы иметь некоторые численные характеристики, количественно описывающие те или иные его свойства. Визуально на картах (рис. 3) можно выделить три области: с преобладанием ядер (зеленого цвета), с преобладанием сосудов (красного цвета) и более темную область, в которой интересующие нас объекты почти полностью отсутствуют. Введем некоторую пороговую величину *thr*, имеющую смысл минимального значения одного из параметров *vess* и *nuclei*, при котором рассматриваемый пиксел карты считается относящимся к области с преобладанием сосудов или ядер. Величина этого порога может выбираться эмпирически или варьироваться. Тогда в качестве численных характеристик изображения образца ткани можно выбрать следующие:

– отношение площади области с преобладанием сосудов к площади пятна (*vess area*);

– отношение площади области с преобладанием ядер к площади пятна (*nuclei area*);

– среднее значение величины *vess* по соответствующей области;

– среднее значение величины *nuclei* по соответствующей области.

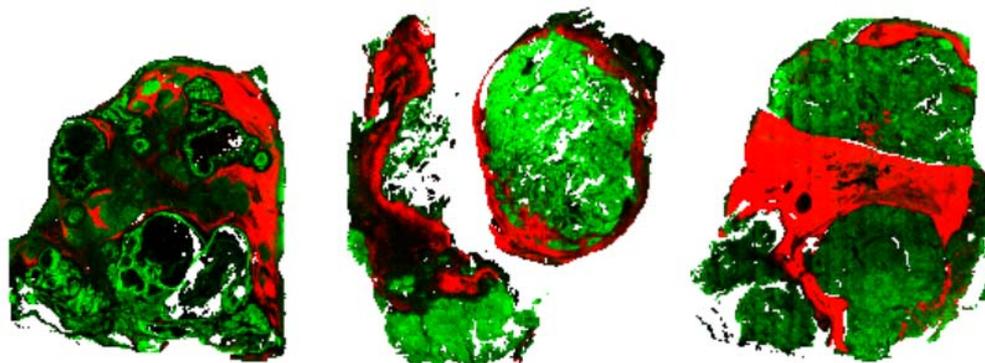


Рис. 3. Примеры цветокодированных карт исследуемых параметров образцов тканей (красный канал соответствует величине *vess*, зеленый канал – величине *nuclei*)

Таким образом, по каждому пациенту были получены четыре численных значения, характеризующие морфологическую структуру изображения раковой ткани, которые в дальней-

шем могут быть использованы при проведении корреляционного анализа с целью выявления связей с имеющимися клиническими данными. Основным достоинством выбранных параметров является то, что они могут быть получены с применением быстрых алгоритмов и в то же время представляют собой достаточно точные оценки реальных параметров. Это обусловлено используемой статистической процедурой коррекции, которая основана на уравнении регрессии, полученном на стадии обучения. Кроме того, весьма важен и тот факт, что эти параметры интуитивно понятны для эксперта-патолога и могут быть им легко проинтерпретированы.

Заключение

В ходе работы был разработан алгоритм картирования гистологических изображений сверхбольшого разрешения. Он основан на идее разбиения сверхбольшого изображения на малые участки, независимой обработки каждой части с извлечением численных характеристик и составления карты распределения найденных характеристик по всему образцу ткани. Полученная карта обладает достаточно малым разрешением и дает возможность визуального анализа морфологической структуры исследуемого образца ткани.

В результате проведенных экспериментальных исследований были картированы характеристики гистологических снимков сверхбольшого разрешения 68 пациентов и по картам были вычислены интересующие нас величины. В ходе работы были выбраны следующие пороги по составляющим цветовой схемы HSV для выделения характерных объектов на изображении:

- для сосудов: (*vess*) *hue* [0; 0,4], *saturation* [0,3; 1], *value* [0; 0,65];
- для ядер: (*nuclei*) *hue* [0,6; 0,76], *saturation* [0,3; 0,6], *value* [0,25; 0,7].

Пороговая величина *thr*, имеющая смысл минимального значения одного из параметров *vess* и *nuclei*, при котором рассматриваемый пиксел карты считается относящимся к области с преобладанием сосудов или ядер, принимала различные значения и для всех значений строились матрицы корреляций между клиническими данными и вычисленными значениями *vess*, *nuclei*, *vess area*, *nuclei area*. Из всех получившихся корреляций отбирались наиболее устойчивые, т. е. сохраняющие свою статистическую значимость при многих различных значениях параметра *thr*. Всего было выявлено 17 подобных корреляций, имеющих уровень значимости $p < 0,05$. Само же значение порога *thr* в результате было выбрано таким, чтобы отобранные корреляции имели наибольшую значимость, и оказалось $thr = 0,15$.

Было установлено, что клиническими данными, значимо коррелирующими с параметрами, вычисленными по результирующим картам образцов раковой ткани, являются: продолжительность жизни с момента начала лечения до момента последнего контакта (статистическая значимость $p < 0,00014$), количество сеансов химиотерапии до ($p < 0,004$) и после ($p < 0,018$) операции, линейные размеры яичников ($p < 0,001$), наличие увеличения живота ($p < 0,010$) и др. Детальное рассмотрение всех значимых корреляций показало, что некоторые из них могут быть легко объяснены на основе имеющихся знаний о характере протекания исследуемого заболевания, в то время как другие потенциально представляют определенный интерес с научной и с практической точек зрения.

Данная работа выполнена в рамках проекта МНТЦ В-1682.

Список литературы

1. Schwab, M. Encyclopedia of Cancer. 2 ed. / M. Schwab. – Heidelberg : Springer, 2009. – 3235 p.
2. Hayat, M. Methods of Cancer Diagnosis, Therapy and Prognosis / M. Hayat. – Heidelberg : Springer, 2009–2010. – 2600 p.
3. Wootton, R. Image Analysis in Histology: Conventional and Confocal Microscopy / R. Wootton, D. Springall, J. Polak. – Cambridge : Cambridge University Press, 1995. – 425 p.
4. Histopathological image analysis: A review / M.N. Gurcan [et al.] // IEEE Reviews in Biomedical Engineering. – 2009. – Vol. 1. – P. 147–171.
5. Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading / O. Sertel [et al.] // Journal of Signal Processing Systems. – 2009. – Vol. 55, № 1. – P. 169–183.

6. Yu, F. Semantic content analysis and annotation of histological images / F. Yu, H. Ip // *Computers in Biology and Medicine*. – 2008. – Vol. 38. – P. 635–649.
7. Critical comparison of 31 commercially available digital slide systems in pathology / M.G. Rojo [et al.] // *International Journal of Surgical Pathology*. – 2006. – Vol. 14. – P. 285–305.
8. A computer image analysis system for microvessel density measurement in solid tumours / J.C. Goddard [et al.] // *Angiogenesis*. – 2002. – Vol. 5. – P. 15–20.
9. Merritt, W.M. Markers of angiogenesis in ovarian cancer / W.M. Merritt, A.K. Sood // *Dis Markers*. – 2007. – Vol. 23. – P. 5–6.
10. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer / S. Petushi [et al.] // *BMC Medical Imaging*. – 2006. – Vol. 6. – P. 6–14.
11. An Imaging Workflow for Characterizing Phenotypical Change in Terabyte Sized Mouse Model Datasets / K. Mosaliganti [et al.] // *Journal of Biomedical Informatics*. – 2008. – Vol. 41. – P. 863–873.
12. Interactive Histology of Large-Scale Biomedical Image Stacks / Jeong Won-Ki [et al.] // *IEEE Transactions on Visualization and Computer Graphics*. – 2010. – Vol. 16, № 6. – P. 1386–1395.
13. Goffe, R. Tiled Top-Down Pyramids and Segmentation of Large Histological Images / R. Goffe, L. Brun, G. Damiand // *Lecture Notes in Computer Science*. – Springer, 2011. – Vol. 6658. – P. 255–264.

Поступила 10.10.11

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, ул. Сурганова, 6
e-mail: {vassili.kovalev, vitali.liauchuk}@gmail.com*

V.A. Kovalev, V.A. Liauchuk

FEATURE MAPPING OF HYPER-LARGE HISTOLOGICAL IMAGES OF CANCEROUS TISSUE

The paper is dedicated to the problem of histological image analysis and it has rather engineering character. A method for mapping features of hyper-large images of cancerous tissue is presented. Being computed, the resultant maps can be employed for both visual and formal quantitative analysis. Statistical analyzes were performed to reveal significant correlations between the output morphological features of the tissue and patients' data.