

АНАЛИЗ ЛЮМИНЕСЦЕНТНЫХ ИЗОБРАЖЕНИЙ ОБЪЕКТОВ МИКРОБИОЛОГИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

Е.В. Лисица, Н.Н. Яцков, В.В. Апанасович

Кафедра системного анализа и компьютерного моделирования, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: {lisitsa, yatskou}@bsu.by

В работе исследуется применение метода главных компонент для увеличения эффективности анализа трехканальных люминесцентных изображений объектов микробиологии на примере раковых клеток тканей опухолей молочной железы.

ВВЕДЕНИЕ

Широкое распространение для медико-биологических исследований получил метод люминесцентной микроскопии [1-3]. Существующие методы люминесцентной микроскопии ориентированы на использование более чем одного цветового маркера, что позволяет проводить более точный анализ исследуемых образцов [4]. С другой стороны использование нескольких красителей приводит к перекрытию их спектров при накачивании образцов и испускании ими флуоресценции. Т.е., каждый цветовой канал <<скрыто>> влияет на интенсивность, получаемую в других канал. Данную проблему можно решить применив в ходе анализа данных один из методов снижения размерности данных, например – метод главных компонент.

Целью данной работы является разработка метода восстановления информации об интенсивности объектов многоканальных люминесцентных изображений с использованием метода главных компонент на примере анализа трехканальных изображений раковых клеток.

I. МАТЕРИАЛЫ И МЕТОДЫ

В работе рассматриваются микрочипы срезов тканей опухолей молочной железы [4]. Пример исходного изображения в градации серого цвета представлен на рис.1А, Б. Экспериментальная часть исследований проведена по методикам, представленным в работе [5, 6]. Изображениями являются популяции клеток, окрашенные в зеленые, синие и красные цвета (трехканальные люминесцентные сигналы в системе RGB). В цитоплазмах раковых клеток регистрируются процессы с участием белка цитокератина [5]. Белок маркируется цианиновым красителем Cy3 и регистрируется в зеленом цветовом канале изображения. Красный канал изображения зарезервирован для индикации ядер раковых клеток. В ядрах раковых клеток находится белок эстроген-рецептор, для маркировки которого используется краситель. Для маркировки ядер используется краситель 4,6-диамидино-2-фенилиндол дигидрохлорид (DAPI, 4',6-diamidino-2-phenylindole) и заре-

зервирован синий канал. Размер изображений – 2048 на 2048 пикселей в каждом из трех каналов, разрешающая способность – 0.2 мкм на пиксель.

II. ТЕОРЕТИЧЕСКОЕ ОПИСАНИЕ

Цветные RGB изображения можно представить в виде трехмерного массива цветных пикселей (рис. 1А, Б, В). Причем каждый пиксель является триплетом, элементы которого соответствуют трем цветовым каналам: зеленому, красному и синему. Тогда интенсивность исходного трёхцветного изображения в псевдоцветах можно представить как сумму интенсивностей трех цветных каналов $I=R+G+B$ [11], где R – красная компонента, G – зелёная компонента, B – синяя компонента, I – итоговое изображение в псевдоцветах.

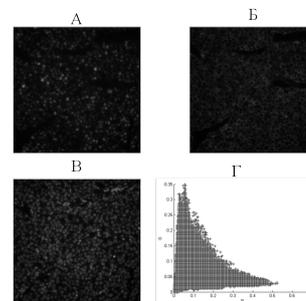


Рис. 1 – Пример полутоновых компонент люминесцентного изображения А) Красная компонента R. Б) Зеленая компонента G. В) Синяя компонента B. Г) График разброса значений интенсивности для красной и зеленой компонент

Зелёный канал G используется для регистрации сигналов красителей из областей цитоплазм раковых клеток. Синий канал B зарезервирован для регистрации сигналов в области ядер. В общем случае, в красном канале регистрируется сигнал красителя биомаркера рака R в ядрах клеток. Между регистрационными каналами изображения наблюдаются корреляции (см. рис. 1Г). Учет корреляции между каналами регистрации является дополнительным источником информации для последующего анализа, что позволит тем самым повысить эффективность анализа изображений. В первом приближении дополнительная информация о расположе-

нии границ цитоплазм клеток может быть восстановлена с применением метода главных компонент (МГК).

Изначально каждый пиксель трехканального изображения с координатами i, j , где $1 \leq i \leq w$ и $1 \leq j \leq h$, w – ширина изображения в пикселях, h – высота изображения, описывается тремя цветовыми компонентами $R(i, j)$ – красная компонента изображения, $G(i, j)$ – зеленая компонента изображения, $B(i, j)$ – синяя компонента изображения.

Представим имеющиеся на изображении пиксели в виде таблицы 1. Каждый признак изображения n_1 описывается двенадцатью признаками X_1, X_2, \dots, X_{12} . Используя МГК [12] найдем соответствующие главные компоненты Z_1, Z_2, \dots, Z_{12} . Поскольку основная задача МГК – выделение существенной информации, то ожидается, что на первые две компоненты будет приходиться большая часть информации необходимой для описания системы. А в последних компонентах – шум с изображения.

III. РЕЗУЛЬТАТЫ

Выполнен анализ 9 изображений. Доли разброса (значения изменяются от 0 до 1), приходящиеся на несколько первых главных компонент – $\delta_1 = 0.58 \pm 0.02$, $\delta_2 = 0.26 \pm 0.02$, $\delta_3 = 0.16 \pm 0.02$. Вклад остальных компонент стремится к 0. Как и ожидалось основной вклад в описание системы вносят первые три компонента (рис. 2), а последние компоненты наименее информативны и содержат шум (рис. 2).

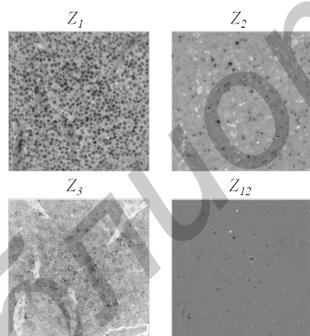


Рис. 2 – Главные компоненты

Основной вклад приходится на первые две главные компоненты. Начиная с 4-ой компоненты вклады вносимые компонентами не значи-

тельные, что позволяет исключить их из дальнейшего анализа и использовать это свойство для фильтрации изображений. Следует отметить, что наличие большого количества выбросов на люминесцентных изображениях может значительно снизить эффективность применения метода главных компонент.

Полученные изображения в координатах главных компонент можно интерпретировать следующим образом – первая компонента содержит информацию о ядрах клеток, во вторая компонента содержит информация о границе ядер. Третья компонента позволяет выделить цитоплазмы клеток. На последующих этапах анализа выполняется сегментация объектов изображений в координатах первых главных компонент, что позволяет значительно повысить эффективность проводимого анализа в сравнении с традиционным подходом на основе независимого анализа каждого из каналов исходного изображения.

IV. ВЫВОДЫ

В работе показана принципиальная возможность использования метода главных компонент для восстановления информации об объектах трехканальных люминесцентных изображений, что позволяет значительно повысить эффективность применения методов сегментации объектов изображений.

1. Ronneberger, O. Spatial quantitative analysis of fluorescently labeled nuclear structures: Problems, methods, pitfalls / O. Ronneberger, D. Baddeley, F. Scheipl, P. J. Verveer, H. Burkhardt, C. Cremer, L. Fahrmeir, T. Cremer and B. Joffe // Chromosome Research.– 2008. – Vol. 16, – P. 523–562.
2. Al-Ali, H., High Content Screening with Primary Neurons, in Assay Guidance Manual, /Al-Ali H., G.S. Sittampalam. – 2004.
3. Абламейко С.В. Медицинские информационные системы и технологии / С. В. Абламейко, В. В. Анищенко, В. А. Лапичкий, А. В. Тузиков // Издательство: ОИПИ НАН Беларуси,- Минск, 2007. - 176 с.
4. Bauer, M. Spot identification and quality control in cell-based microarrays. / M. Bauer et all // ACS Comb Sci.– 2012. –Vol.14, –P. 471–7.
5. Chung, G.G., Quantitative analysis of estrogen receptor heterogeneity in breast cancer /G. G. Chung et all // Lab Invest.– 2007. – Vol.87, –P. 662-9.
6. R.L. Camp Automated subcellular localization and quantification of protein expression in tissue microarrays / R. L. Camp, G. G. Chung and D.L. Rimm // Nat. Med.–2002. –Vol.8, –P. 1323-7.

Таблица 1 – Набор многомерных данных, представляющий собой пиксели изображения n_1, n_2, \dots, n_{i*j}

N	$X_1 = R_{ij}$	$X_2 = G_{ij}$	$X_3 = B_{ij}$	$X_4 = X_1^{-1}$	$X_5 = X_2^{-1}$	$X_6 = X_3^{-1}$	$X_7 = X_2 * X_4$	$X_8 = X_3 * X_4$	$X_9 = X_7^{-1}$	$X_{10} = X_8^{-1}$	$X_{11} = X_2 * X_6$	$X_{12} = X_{11}^{-1}$
n_1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	$x_{1,5}$	$x_{1,6}$	$x_{1,7}$	$x_{1,8}$	$x_{1,9}$	$x_{1,10}$	$x_{1,11}$	$x_{1,12}$
n_2	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	$x_{2,5}$	$x_{2,6}$	$x_{2,7}$	$x_{2,8}$	$x_{2,9}$	$x_{2,10}$	$x_{2,11}$	$x_{2,12}$
...
n_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}	x_{i*j}