

Выводы

1. Рассмотрены приемы повышения производительности искусственных нейронных сетей в задачах медицинской диагностики при реализации параллельных алгоритмов с применением CUDA-устройств.

СПИСОК ЛИТЕРАТУРЫ

1. Галушкин А.И. О методике решения задач в нейросетевом логическом базисе // Нейрокомпьютер. – 2006. – № 2. – С. 49–70.
2. Мызников А.В., Россиев Д.А., Лохман В.Ф. Нейросетевая экспертная система для оптимизации лечения облитерирующего тромбангиита и прогнозирования его непосредственных исходов // Ангиология и сосудистая хирургия. – 1995. – № 2. – С. 100–103.
3. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. – Новосибирск: Наука, 1996. – 276 с.
4. Harris M. Mapping Computational Concepts to GPUs // GPU Gems 2. – 2006. – № 2. – P. 493–508.
5. Hall J.D., Carr N.A., Hart J.C. Cache and Bandwidth Aware Matrix Multiplication on the GPU. Technical Report. – UIUCDCS-R-2003–2328. 2012. URL: <http://graphics.cs.uiuc.edu/~jch/papers/UIUCDCS-R-2003–2328.pdf> (дата обращения: 17.01.2012).
6. Horn D. Stream Reduction Operations for GPGPU Applications // GPU Gems 2. – Addison Wesley, 2006. – P. 573–589.
7. Хайкин С. Нейронные сети полный курс. – М.: ООО «И.Д. Вильямс», 2006. – 1104 с.
8. Rossiev D.A., Savchenko A.A., Borisov A.G., Kochenov D.A. The employment of neural-network classifier for diagnostics of different phases of immunodeficiency // Modelling, Measurement & Control. – 1994. – V. 42. – № 2. – P. 55–63.
9. Горбань А.Н., Дунин-Барковский В.Л., Кардин А.Н. Нейроинформатика / Отв. ред. Е.А. Новиков. – Новосибирск: Наука, 1998. – 296 с.
10. NVIDIA CUDA Programming Guide Version 2.3.1 // NVIDIA – World Leader in Visual Computing Technologies. 2011. URL: http://www.nvidia.com/object/cuda_develop.html (дата обращения: 17.01.2012).

Поступила 24.01.2012 г.

УДК 004.652.3

ПРИМЕНЕНИЕ МЕТОДОВ ВИЗУАЛИЗАЦИИ ПРИ ИССЛЕДОВАНИИ СТРУКТУРЫ ЭКСПЕРИМЕНТАЛЬНЫХ МНОГОМЕРНЫХ ДАННЫХ

В.А. Воловоденко, О.Г. Берестнева, Е.В. Немеров*, И.А. Осадчая

Томский политехнический университет

*Сибирский государственный медицинский университет, г. Томск

E-mail: ogb@tpu.ru

Рассмотрены методы структурного анализа многомерных данных, представлены различные подходы к визуализации результатов экспериментальных исследований. Приведены примеры решения прикладных задач с использованием NovoSparkVisualizer.

Ключевые слова:

Кластерный анализ, методы визуализации, когнитивная графика, спектральные представления.

Key words:

Cluster analysis, imaging, cognitive graphics, spectral representation.

В настоящее время накоплен обширный арсенал средств анализа многомерных данных. Наиболее полное изложение применяемых здесь подходов, сопровождающееся подробными ссылками на ключевые работы, содержится в [1]. В [2] приведена классификация основных методов анализа структуры многомерных данных:

- 1) визуализация данных: линейные методы снижения размерности, нелинейные отображения, многомерное шкалирование, заполняющие пространство кривые;
- 2) автоматическое группирование: факторный и кластерный анализ объектов и признаков, иерархическое группирование, определение «точек сгущения».

В основу приведенной классификации положен признак, отображающий степень участия экспериментатора в выделении особенностей взаимоотношений между исследуемыми объектами и признаками. Применение методов визуализации данных нацелено на поиск наиболее выразительных изображений совокупности исследуемых объектов для последующего максимального задействования потенциала зрительного анализатора экспериментатора.

Компьютерная обработка данных предполагает некоторое математическое преобразование данных с помощью определенных программных средств. Для этого необходимо иметь представление как о математических методах обработки данных, так и о соответствующих программных средствах [2].

Целью работы являлось стремление визуализировать образы, представленные численными данными, и показать возможность выявления отношений между ними на основе наблюдений за совокупностью полученных образов.

Применение предложенного метода, на наш взгляд, приводит к значительному упрощению процедуры визуализации и способствует выявлению скрытых связей в очевидной форме. Это позволяет не только использовать свойства метода, но и способствует учету опыта группы экспертов, рассматривающих конкретную проблему. Последнее обстоятельство повышает актуальность предложенного подхода.

Методы визуализации позволяют исследователю одним взглядом обнаружить особенности, выявить закономерности и аномалии в больших объемах информации. Основной задачей визуализации данных является задача получения визуального образа, однозначно соответствующего набору данных [3]. В данной статье рассмотрен подход [4, 5], который позволяет провести визуализацию основных линейных конструкций: отрезок, ломаная линия, симплекс в многомерных пространствах. Основой визуализационного подхода является линейное преобразование значений многомерного наблюдения A в двумерную кривую $f_A(t)$, т. е. $A \leftrightarrow f_A(t)$, при этом гарантируется, что близким по значениям наблюдениям A и B будут соответствовать визуально близкие образы-кривые $f_A(t)$ и $f_B(t)$; для сильно различающихся по значениям наблюдений их образы-кривые будут заметно отличаться.

В рассматриваемом случае наиболее общей формой представления данных является вектор конечномерного пространства R_n

$$A = (a_0, a_1, a_2, \dots, a_{n-1}) \in R_n. \quad (1)$$

Для перехода от данного вектора к визуальному образу будет использоваться базис ортонормированных функций $\{\varphi_i(\tau)\}_{i=0}^{\infty}$. В качестве такого базиса можно использовать известные функции, в частности ортонормированные полиномы Лежандра на отрезке $[0, 1]$, множество которых обозначим через $\{l_i(\tau)\}_{i=0}^{\infty}$.

В таком случае точке с координатами $A=(a_0, a_1, a_2, \dots, a_{n-1})$ можно поставить в соответствие функцию:

$$F_A(\tau) = \sum_{i=0}^{n-1} a_i l_i(\tau). \quad (2)$$

Формирование вектора A связано с преобразованием данных. Для характеристики наблюдаемого многомерного объекта большую роль играют значения его координат. В большинстве случаев каждый показатель имеет свою единицу измерения, и его значение будет влиять на вид функции $F_A(\tau)$. Для того чтобы исключить влияние разноименности показателей на вид функции $F_A(\tau)$, необходимо перейти к безразмерным единицам одним из известных способов. Следует отметить, что порядок

включения показателей в вектор A , также будет влиять на вид функции $F_A(\tau)$. Разница между формулами (1) и (2) заключается в том, что для вектора A из (1) возможно только аналитическое представление, в то время как для функции $F_A(\tau)$ возможно представление в виде графика этой функции. Между (1) и (2) устанавливается однозначная связь в обе стороны, т. е. взаимно-однозначная связь. Если ввести в рассмотрение второй вектор

$$B = (b_0, b_1, b_2, \dots, b_{n-1}) \in R_n,$$

то ему ставится в соответствие функция

$$F_B(\tau) = \sum_{i=0}^{n-1} b_i l_i(\tau).$$

Будем считать, что функции $F_A(\tau)$ и $F_B(\tau)$ являются визуальными образами точек A и B , принадлежащих пространству R_n (рис. 1).

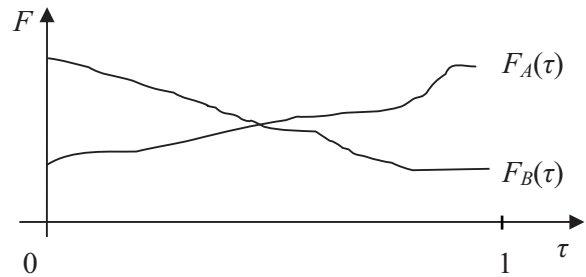


Рис. 1. Визуальные образы точек A и B , принадлежащих пространству R_n

Чем больше кривые неотличимы друг от друга, тем идентичнее наблюдения, которые они представляют, т. е. метод устанавливает взаимно-однозначное соответствие между строками в наборе данных и их кривыми.

Если отобразить кривые наблюдений в трехмерном пространстве, используя третье измерение, называемое также « Z -измерением», как расстояние в многомерном пространстве или промежуток времени между двумя наблюдениями, можно обнаружить много интересных свойств. Ось Z — это ось следования образов. В простейшем случае значения координаты Z совпадает с номером образа, но этой координате можно придать значение расстояния в пространстве признаков от начала координат до образа (объекта).

Представленный подход реализован в пакете *NovoSpark Visualizer* [6], на базе которого авторами был успешно решен ряд прикладных задач анализа и интерпретации многомерных данных в социальной сфере [7] и медицине [5].

Рассмотрим результаты решения двух прикладных задач, иллюстрирующих эффективность представленного метода визуализации многомерных экспериментальных данных.

В качестве исходной информации первой задачи имеем выборочные сведения о беременных женщинах и соответствующие им результаты лабораторных и инструментальных исследований (подробное описание экспериментальных данных представлено в [8–10]). Экспериментальная вы-

борка состоит из трех групп беременных женщин (30 человек) в возрасте от 18 до 44 лет.

Первую (контрольную) группу составили 10 практически здоровых беременных женщин в возрасте от 20 до 42 лет, не получавших оздоровительных мероприятий. Вторую (группу сравнения) составили 10 беременных женщин в возрасте от 18 до 40 лет, имевших соматические заболевания, не получавших комплекса оздоровительных мероприятий. Третью (основную группу) составили 10 беременных женщин в возрасте от 19 до 43 лет, имевших соматические заболевания, которым проводился различный комплекс оздоровительных мероприятий.

Изменение их состояние оценивалось в I и III триместрах беременности по набору некоторых медико-биологических показателей. В табл. 1 представлен фрагмент таблицы исходных данных.

В качестве эталонного набора выбрана контрольная группа. Выберем по одному наблюдению из каждой группы:

- 1 – контрольная;
- 2 – сравнения;
- 3 – основная.

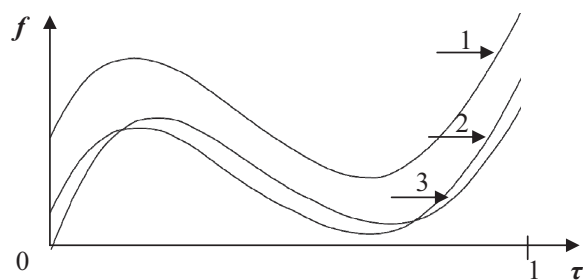


Рис. 2. Сравнение наблюдений I триместра

Как видно из рис. 2, наблюдения основной группы и группы сравнения очень похожи и значительно отличаются от наблюдений контрольной группы, что свидетельствует о том, что исходный набор переменных состояния беременных «подходит» для оценки динамики их состояния при проведении комплекса оздоровительных мероприятий.

Рассмотрим кривые наблюдений в III триместре.

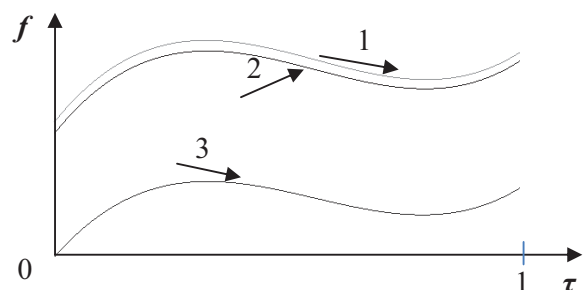


Рис. 3. Сравнение наблюдений III триместра

Из рис. 3 и 4 видно, что наблюдение из основной группы сильно приблизилось к наблюдению контрольной группы, а наблюдение группы сравнения практически не изменилось.

Это означает, что к III триместру при получении комплекса оздоровительных мероприятий показате-

тели беременных женщин, имевших соматические заболевания, значительно приближаются к показателям практически здоровых беременных женщин.

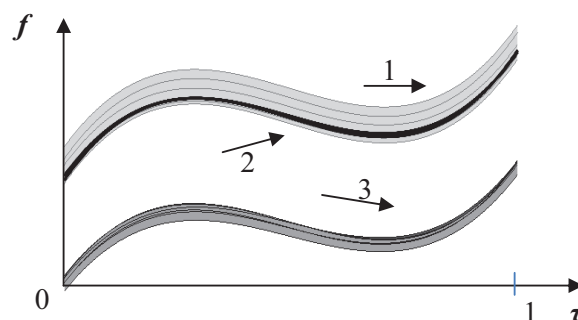


Рис. 4. Сравнение группы эталонов

Рассмотрим реализацию подхода, предложенного в [6], позволяющего отображать многомерные объекты в виде кривых или «спектров». «Спектральные представления» в данном методе подчеркивают отличительные характеристики каждой кривой и помогают более детально исследовать их визуальные свойства. Цветная палитра акцентирует уровни изменения значений кривых. Производя воображаемое упорядочение кривых вдоль оси и, наблюдая сверху на результат этой операции, можно получить цветные полосы, представляющие собой спектр каждого наблюдения [7]. Цветовая гамма определена по аналогии с картой, где наибольшим значениям функции соответствуют красные тона, а наименьшим значениям соответствуют холодные, синие оттенки.

Выберем 4 наблюдения из сравнительной (группа сравнения) и основной групп I триместра. Построим горизонтальные спектры выбранных наблюдений (рис. 5).

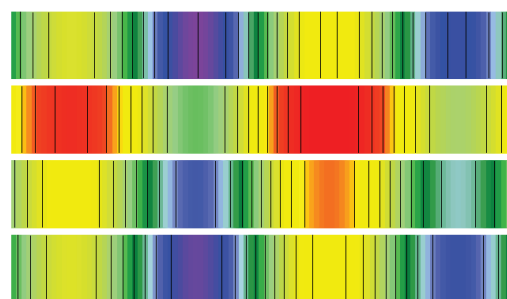


Рис. 5. Спектры выбранных наблюдений I триместра

Из рис. 5 видно, что при сопоставлении полихроматических рядов каждого из 4 горизонтальных спектров, построенных по единой технологии можно найти основания для их объективной классификации и соизмерения. Для этого могут быть использованы многоэтапные (многоуровневые) процедуры сравнения по качественным хроматическим признакам.

На первом этапе таких процедур ищется доминирующий признак различия, применяемый в качестве основания для разбиения всего множества спектров на классы. Таким дифференцирующим признаком может быть Pr_1 – наличие в спектрах

хроматических компонентов наиболее темного цвета (в нашем случае – фиолетового). Тогда, в соответствии с Pr_1 , на первом этапе из исходной совокупности индивидуальных спектров наблюдений $\{1, 2, 3, 4\}$ выделяются классы $K_{11}=\{1, 4\}$ и $K_{12}=\{2, 3\}$. На втором этапе в классе K_{12} используется признак Pr_2 – наличие компонентов менее темного цвета, чем тот, который использовался для формулирования Pr_1 (в данном случае, темно-синий). Данный признак производит разбиение класса K_{12} на два: $K_{21}=\{2\}$ и $K_{22}=\{3\}$. На данном этапе возможности дальнейшего формирования и применения признаков, аналогичных Pr_1 , Pr_2 , необходимых для продолжения классификации исчерпаны. Окончательными результатами проводимой классификации будут классы $K_{11}=\{1, 4\}$, $K_{21}=\{2\}$ и $K_{22}=\{3\}$.

Аналогично разделяем на классы выборку, которая содержит 6 наблюдений из сравнительной (группа сравнения) и основной групп III триместра.

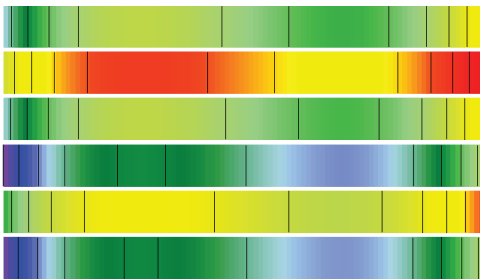


Рис. 6. Спектры выбранных наблюдений III триместра

В результате выделены 4 класса: $K_{11}=\{4, 6\}$, $K_{21}=\{1, 3\}$, $K_{31}=\{5\}$ и $K_{32}=\{2\}$.

Исходная информация второй прикладной задачи представляет собой данные о пациентах с четырьмя типами заболеваний, а именно:

- бронхиальная астма непсихогенная (*BANP*);
- бронхиальная астма сомато-психогенная (*BASP*);
- бронхиальная астма психогенно-индуцированная (*BAPI*);
- психогенная одышка (*PD*).

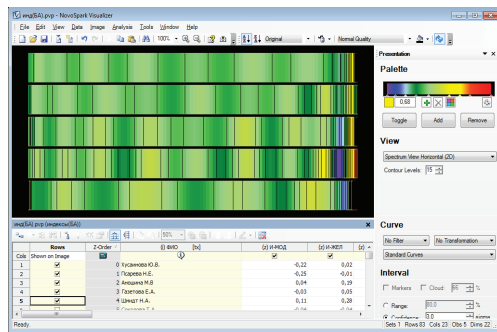
Проведем анализ показателей физиологических реакций бронхолегочной системы в ответ на психофизиологическое воздействие (аудиовизуальную стимуляцию). Технология получения экспериментальных данных подробно изложена в [10].

Сравним выборочные данные для четырех форм бронхиальной астмы по визуальной близости спектров наблюдений (рис. 7).

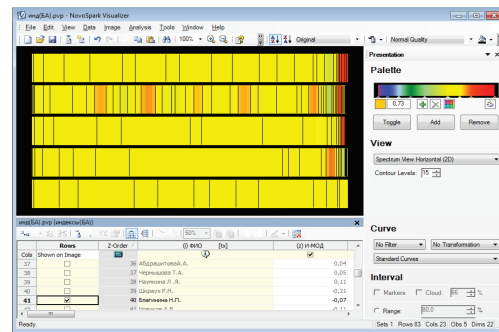
Каждая цветная полоска в спектральном виде соответствует показателю одного пациента. В нашем случае на рис. 7 представлено по пять цветных полосок, соответственно представляющих по пять пациентов с различными формами бронхиальной астмы. Цветовые «спектры» пациентов с диагнозом *BAPI* и *PD* схожи. То же можно сказать и о пациентах с диагнозом *BASP* и *BANP*.

Как видно из рис. 7, наиболее близкие показатели имеют пациенты с *BASP* и *PD*. Наиболее выраженные различия можно заметить у пациентов с *BAPI*.

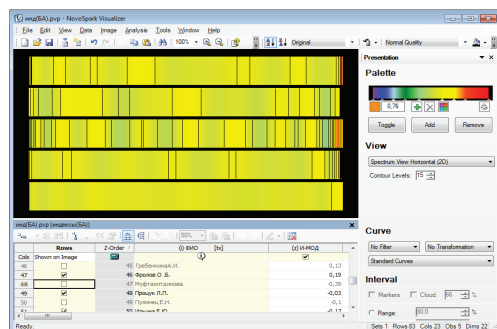
Таким образом, использование средств когнитивной графики позволило выявить некоторые



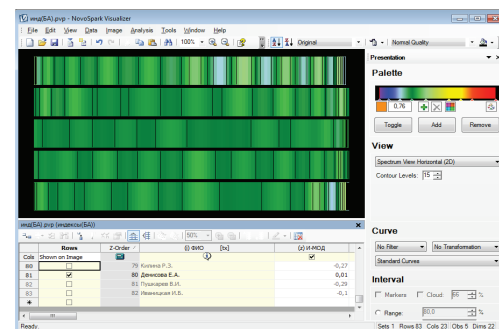
а



б



в



г

Рис. 7. Спектральные представления данных о пациентах с различными диагнозами: а) *BAPI*; б) *BASP*; в) *BANP*; г) *PD*

не известные ранее закономерности физиологических реакций бронхолегочной системы в ответ на психофизиологическое воздействие.

Наиболее ярко отражает черты многомерного анализа в классификации кластерный анализ. Кластер – объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определёнными свойствами. Главное назначение кластерного анализа – разбиение множества исследуемых объектов и признаков на однородные, в соответствующем понимании группы или кластеры. Это означает, что решается задача классификации данных и выявления соответствующей структуры в ней. Методы кластерного анализа можно применять в самых различных случаях, даже в тех случаях, когда речь идет о простой группировке, в которой все сводится к образованию групп по количественному сходству [2]. Основное достоинство кластерного анализа в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. В табл. 1 и 2 представлены результаты разбиения, полученные с помощью кластерного анализа (метод Уорда) на базе пакета *Statistica*.

В табл. 1 представлены результаты кластеризации по физиологическим показателям бронхолегочной системы, отражающим динамику их изменения после проведения сеансов аудиовизуальной стимуляции.

Таблица 1. Распределение пациентов по кластерам (степень воздействия аудиовизуальной стимуляции)

Диагноз	Количество пациентов			
	1 кластер	2 кластер	3 кластер	4 кластер
<i>BANP</i>	1	5	15	8
<i>BASP</i>	3	6	5	4
<i>BAPI</i>	8	14	2	0
<i>PD</i>	1	0	6	5

В 1-м кластере преимущественно оказались пациенты с диагнозом *BAPI*, а также 3 человека с *BASP*, 1 – с *BANP* и 1 – с *PD*.

Во второй кластер преимущественно попали также больные *BAPI*, 6 человек – с диагнозом *BASP* и 5 человек – с *PD*.

Третий кластер объединил преимущественно людей с диагнозом *BANP*. Также сюда попали 2 человека с *BAPI*, 5 человек – с *BASP* и 6 – с *PD*.

В четвертом кластере также преимущественно оказались люди с *BANP*, а также 4 – с *BASP* и 5 – с *PD*.

На основе анализа полученных результатов (значений центроидов для каждого кластера) был сделан вывод о том, что на пациентов 1, 3 и 4 кластеров аудиовизуальная стимуляция мозга оказывает аналогичное воздействие – улучшаются значе-

ния показателей вентиляции легких и механики дыхания. При этом для первого кластера улучшение значений показателей вентиляции легких менее выражено, чем для третьего кластера. Улучшение статической растяжимости легких характерно только для 4 кластера. Для пациентов, попавших во 2 кластер, аудиовизуальная стимуляция практически не оказывает никакого воздействия на изменение показателей вентиляции легких и механики легких.

В табл. 2 представлены результаты кластеризации по исходным («фоновым») физиологическим показателям.

Таблица 2. Распределение пациентов по кластерам («фоновые» показатели бронхолегочной системы)

Диагноз	Количество пациентов			
	1 кластер	2 кластер	3 кластер	4 кластер
<i>BANP</i>	6	21	0	2
<i>BASP</i>	3	6	7	2
<i>BAPI</i>	14	1	7	2
<i>PD</i>	1	8	2	1

В первый кластер попали, в основном, пациенты с диагнозом «психогенно-индуцированная астма», во второй – пациенты с непсихогенной бронхиальной астмой и с психогенной отдышкой, в третий – пациенты с психогенно-индуцированной и сомато-психогенной бронхиальной астмой. И, наконец, в четвертом оказалось практически одинаковое количество представителей каждого из заболеваний. Полученные результаты показывают, что на основе имеющихся показателей бронхолегочной системы не удалось выделить кластеры, соответствующие медицинским диагнозам.

Выводы

Показано, что методы визуализации многомерных данных и подходы на их основе перспективны для анализа и наглядного представления экспериментальных результатов.

Представление *многомерного* наблюдения в виде *двумерного* образа (кривой) гарантирует, что объектам, близким по значениям их характеристик будут соответствовать визуально близкие образы-кривые; для наиболее различающихся – их образы-кривые будут заметно отличаться. Становится возможным автоматически классифицировать наблюдения, определять наиболее важные переменные в модели, производить кластеризацию данных, визуально сравнивать индивидуальные наблюдения и целые наборы данных, а также выполнять много других задач в работе с многомерными данными.

Спектральное представление визуального образа является инструментом, наиболее подчеркивающим различия или сходства образов, чем традиционные методы, характеризующие эти свойства на уровне числовых параметров.

СПИСОК ЛИТЕРАТУРЫ

1. Симчера В.М. Методы многомерного анализа статистических данных. – М.: Финансы и статистика, 2008. – 400 с.
2. Дюк В., Эмануэль В. Информационные технологии в медико-биологических исследованиях. – СПб.: Питер, 2003. – 528 с.
3. Шаропин К.А., Берестнева О.Г., Шкатова Г.И. Визуализация результатов экспериментальных исследований // Известия Томского политехнического университета. – 2010. – Т. 316. – № 5. – С. 172–176.
4. Берестнева О.Г., Пеккер Я.С., Шаропин К.А., Воловоденко В.А. Выявление скрытых закономерностей в медицинских и социально-психологических исследованиях // Аппликативные вычислительные системы: Труды II Междунар. конф. по аппликативным вычислительным системам. – г. Москва, 29–31 октября 2010. – М.: Институт Актуального образования «ЮрИнфоР-МГУ», 2010. – С. 287–296.
5. Шаропин К.А., Берестнева О.Г., Воловоденко В.А., Марухина О.В. Визуализация медицинских данных на базе пакета NovoSpark // Известия Южного федерального университета. Сер. Технические науки. – 2010. – Т. 109. – № 8. – С. 242–249.
6. Воловоденко В.А., Эйдензон Д.В. Визуализация и анализ многомерных данных с использованием пакета «NovoSparkVisualiz-
er». 2008. URL: <http://www.tsu.ru/storage/iro/k020410/s4/s4.doc> (дата обращения: 25.11.2011).
7. Марухина О.В., Берестнева О.Г., Шаропин К.А., Осадчая И.А. Когнитивная графика в социально-психологических исследованиях // Информационные и математические технологии в науке и управлении: Труды XVI Байкальской Всерос. конф. – г. Иркутск, 2011. – С. 176–181.
8. Кривоногова Т.С., Желев В.А., Тропова Т.Е., Гергет О.М. Роль раннего комплекса оздоровительных мероприятий в охране здоровья матерей и их детей // Российский вестник перинатологии и педиатрии. – 2009. – Т. 54. – № 3. – С. 14–19.
9. Гергет О.М., Берестнева О.Г. Интеллектуальная система оценки состояния здоровья беременных женщин и детей // Информатика и системы управления. – 2010. – № 2. – С. 61–64.
10. Немеров Е.В., Языков К.Г. К вопросу изучения личностных свойств в психофизиологической реактивности больных бронхиальной астмой на аудиовизуальную стимуляцию // Вестник Томского государственного педагогического университета. – 2011. – Вып. 6 (108). – С. 134–137.

Поступила 06.04.2012 г.

УДК 004.9

ОЦЕНИВАНИЕ СОСТАВА ИЗОБРАЖЕНИЙ КЛЕТОЧНЫХ ПРЕПАРАТОВ ДЛЯ МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

М.С. Тарков, Н.В. Тихонов, В.Г. Половинкин

Институт физики полупроводников им. А.В. Ржанова СО РАН, г. Новосибирск
E-mail: tarkov@isp.nsc.ru; niknsk@list.ru; pvg@isp.nsc.ru

Предложен метод оценивания состава изображений, основанный на применении полутоновой эрозии и дилатации, преобразования «верх шляпы» и последующей пороговой сегментации. Метод реализован в виде программного продукта и используется в рамках методики цейтраферной (замедленной) компьютерной видеосъемки микроскопических объектов, разработанной для изучения барьерных свойств мембран растительных клеток. Он позволяет оценить относительную площадь наблюдаемых объектов на изображении и может быть использован в различных приложениях, когда требуется определить скорость распада клеточных структур.

Ключевые слова:

Оценивание состава изображений, клеточные препараты.

Key words:

Estimation of the image, cell preparations.

Введение

В настоящее время существует большое количество программных пакетов, предназначенных для анализа биомедицинских изображений, в том числе изображений цитологических препаратов [1, 2]. В то же время в используемых на практике программах анализ ряда характеристик клеток (периметр, площадь, количество) недостаточно автоматизирован. Данная работа ставит своей целью ликвидацию указанного недостатка существующих программных средств.

Изучение свойств мембран клеток в ответ на внешние воздействия – одна из ключевых задач современной биологии и медицины. Известно, что неблагоприятные условия окружающей среды со-

провожают нарушениями устойчивости клеточных мембран и могут приводить к развитию различных форм стресса [3]. Вот почему важно иметь технологию для оперативной оценки барьерных свойств мембран. Один из подходов к такой оценке связан с анализом изображений клеточных препаратов. При неблагоприятных внешних воздействиях (стресс) происходит разрушение клеточных мембран и гибель клеток. В результате количество клеток на изображении уменьшается (рис. 1). По динамике изменения количества клеток на серии изображений, полученной цейтраферной (замедленной) компьютерной видеосъемкой (ЦКВ) [4], можно судить об устойчивости клеточных мембран к внешним воздействиям.