

МЕТОД ДЕКАРТОВА ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ ДЛЯ АНАЛИЗА ИЗОБРАЖЕНИЙ РАЗВИВАЮЩЕГОСЯ ГЛАЗА МУШКИ ДРОЗОФИЛЫ

© 2023 г. Н.А. Данилов*, К.Н. Козлов*, С.Ю. Суркова*^{*,#}, М.Г. Самсонова*

**Санкт-Петербургский политехнический университет Петра Великого,
Политехническая ул., 29, С.-Петербург, 195251, Россия*

#E-mail: surkova_syu@spbstu.ru

Поступила в редакцию 16.12.2022 г.

После доработки 16.12.2022 г.

Принята к публикации 21.12.2022 г.

Методы автоматического выделения признаков привлекают все большее внимание при решении современных задач обработки изображений. Конфокальные изображения однослойного эпителия развивающегося глаза плодовой мушки дрозофилы представляют собой удобную модельную систему для разработки методов выделения сложных признаков. Целью данной работы было применение метода декартова генетического программирования для выявления границ омматидиев – светочувствительных единиц будущего глаза. Использование декартова генетического программирования для анализа картин экспрессии маркера Fasciclin III показало хорошие результаты. Это дает интересные перспективы для дальнейшего применения этой технологии с целью автоматического анализа изображений, полученных с помощью конфокальной микроскопии.

Ключевые слова: обработка изображений, метод декартова генетического программирования, глазной диск, дрозофила

DOI: 10.31857/S0006302923030195, EDN: FTCIPG

Обработка изображений эпителиальных слоев имеет ключевое значение для научных исследований и неинвазивных методов диагностики в медицине [1]. Важную роль играет точность выявления границ клеток эпителия для количественного определения уровней экспрессии генов в каждой клетке.

Удобной модельной системой для разработки методов сегментации изображений эпителиальных тканей является однослойный эпителий развивающегося глаза плодовой мушки дрозофилы (глазного диска) [2]. Эта структура формируется у личинок третьего возраста и включает две зоны. Зона пролиферации состоит из недифференцированных клеток, а зона дифференцировки представляет собой сетчатую структуру омматидиев – презумптивных светочувствительных единиц будущего глаза [3, 4]. Дифференцировка осуществ-

ляется путем прохождения морфогенетической борозды в постериорно-антериорном направлении глазного диска [5–7]. Особенности данной системы является гетерогенность форм и размеров клеток, а также их взаимного расположения. Помимо этого, однослойный эпителий глазного диска образует выпуклость, что приводит к загибам и деформациям клеточного слоя на препаратах. При получении трехмерного набора оптических срезов на конфокальном микроскопе это проявляется в искажении реальной клеточной структуры. Такие особенности этой ткани позволяют разрабатывать методы, корректирующие дефекты биологических образцов и применимые для сегментации изображений сложных объектов.

Целью данной работы было применение метода декартова генетического программирования (ДГП) для обработки изображений с целью выявления границ омматидиев в зоне дифференци-

Сокращение: ДГП – декартово генетическое программирование.

ровки глазного диска. Так как каждый омматидий состоит из клеток-предшественников фоторецепторов, при флуоресцентном окрашивании границы индивидуальных клеток мешают выявлению границ всего омматидия.

Выделение признаков является важной темой в машинном обучении и становится ключевой задачей при работе с необработанными многомерными данными, которые не имеют четко определенных признаков [8–10]. Ручная идентификация в таких случаях часто невозможна, поскольку требует больших объемов ручного труда, времени, а также знаний в предметной области [11, 12]. Современные методы автоматического выделения признаков привлекают все большее внимание при решении современных задач обработки изображений. Большинство таких методов реализовано в виде многоуровневых нейронных сетей. Однако рассматриваемый в данной работе метод ДГП имеет два важных преимущества: 1) часто не требует больших обучающих наборов для изучения конкурентных моделей; 2) как правило, устойчив к зашумленным данным. Принцип метода базируется на предположении о том, что наиболее значимые высокоуровневые признаки могут быть закодированы композицией простейших функций. Метод ДГП был впервые предложен в 1998 г. для создания электрических схем [13]. В последнее время он получил широкое распространение для обработки изображений. В литературе описано всего несколько примеров применения ДГП для выделения признаков из двумерных изображений, где высокоуровневые признаки выделяют напрямую из низкоуровневых [10, 14–16]. В данной работе метод ДГП впервые реализован для обработки изображений развивающегося глаза, что дает интересные перспективы для дальнейшего применения этой технологии для автоматического анализа конфокальных изображений.

ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ

Исходные данные представляют собой трехмерные наборы изображений, полученных с помощью конфокального микроскопа [2] и имеющих размер 2048×2048 пикселей.

Флуоресцентное окрашивание белка Fasciclin III, являющегося маркером латеральных клеточных стенок [17], проводили у личинок *Drosophila melanogaster* третьего возраста методом непрямой

иммунофлуоресценции с помощью первичных антител anti-Fasciclin III (Developmental Studies Hybridoma Bank, NIH, США) и вторичных антител Alexa Fluor 405 (Invitrogen, США) [2].

ПОСТРОЕНИЕ АНАЛИТИЧЕСКОЙ ФУНКЦИИ МОДЕЛИ

Большинство данных высокой размерности, встречающихся в природе, демонстрирует корреляцию между признаками низкого уровня, выражаемыми дополнительными измерениями. Положение пикселей относительно друг друга в матрице изображения чрезвычайно важно для выражения соответствующих высокоуровневых характеристик.

В работе [10] был предложен метод для извлечения признаков из двумерных изображений на основе значений и взаимном расположении пикселей. С помощью ДГП создается модель в виде аналитической функции f такой, что для пикселя p с координатами $(c1, c2)$ в изображении $f(c1, c2) = p$. Модель представляется в виде дерева:

- с нетерминальными символами – алгебраические операции (+, −, ×, безопасное деление, Min и Max) и тригонометрические операции (sin и cos);

- с терминальными символами – переменные (координаты $c1, c2$) и десятичные константы в интервале $(-1; 1)$.

Свойство непрерывности аналитической функции устанавливает отношения между соседними низкоуровневыми функциями (пикселями). Однако поскольку безопасное деление, Min и Max вносят разрыв, модели также могут кодировать пространственные отношения, которые трудно смоделировать в одной непрерывной функции.

ПАРАМЕТРИЗАЦИЯ МОДЕЛИ

Ключевой проблемой использования пространственной информации для выделения признаков является их изменчивость в изображениях, которые представляют одну и ту же концепцию. Например, при классификации рукописных цифр «1» (с засечками) или «|» (ровный штрих без засечек, аналог латинской строчной буквы l [эл] или прописной I [ай]) являются двумя популярными стилями написания одного и того же числа. У этих стилей разные, хотя и похожие пространственные отношения.

Одна и та же функция не может выводить разные значения пикселей для одной и той же координаты и, следовательно, не может кодировать оба стиля. Для решения этой задачи функция

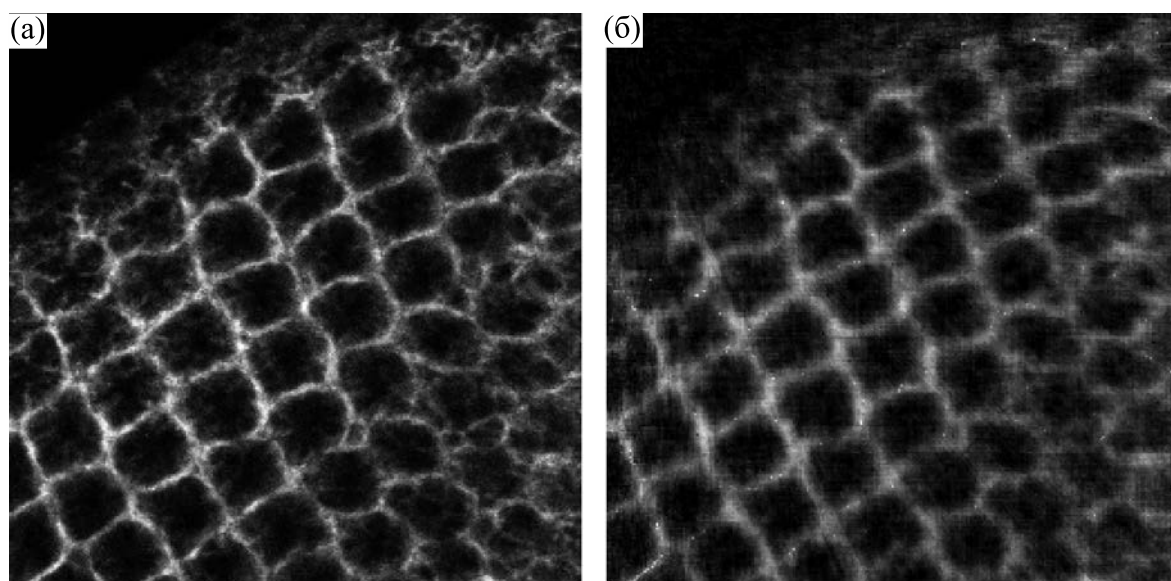


Рис. 1. (а) — Исходное конфокальное изображение участка области дифференцировки глазного диска дрозофилы, окрашенное маркером на Fasciclin III. Светочувствительные единицы (омматидии) представляют собой сетчатую структуру. Частичное окрашивание клеток-предшественников фоторецепторов внутри каждого омматидия создает сложности при выявлении границ омматидиев. (б) — Изображение, восстановленное в течение 6000 внешних итераций.

модели f представляется в виде линейной комбинации нескольких созданных ДГП-моделей, называемых частичными, каждая из которых фокусируется на определенном высокоуровневом признаке в изображении. Параметры линейной комбинации подбираются для наилучшего описания конкретного изображения.

Критерием приспособленности модели является среднеквадратическая ошибка, т. е. средний квадрат разности между предсказанным и фактическим значением пикселя. Таким образом, на начальном этапе вычислений метод фокусируется на тех элементах изображения, которые приводят к большим ошибкам. Каждая следующая частичная модель описывает разность исходного изображения и комбинации предыдущих моделей, таким образом, следующая итерация будет сосредоточена на характеристиках изображений, которые на предыдущей итерации не были хорошо аппроксимированы. Чем ниже среднеквадратическая ошибка, тем лучше приспособлена модель.

ВЫДЕЛЕНИЕ ГРАНИЦ ОММАТИДИЕВ

Разработанный подход был применен к имеющимся изображениям паттернов экспрессии Fasciclin III в развивающемся глазе дрозофилы. В результате построения линейной комбинации 6000 частичных моделей были получены сглаженные

изображения с незначительными артефактами на полностью черном фоне (рис. 1). При этом исключение из линейной комбинации части моделей дало возможность более четко выделить черты, необходимые для дальнейшей обработки.

Наиболее интересными оказались результаты при исключении первых 500 и 1000 частичных моделей (рис. 2). Благодаря использованию аналитической функции модели для определения границ был применен метод вычисления модуля градиента с помощью конечных разностей с шагом 10^{-3} , подобранным опытным путем. Границы, выявленные с помощью вычисления модуля градиента, представлены на рис. 3.

Далее к восстановленному изображению и изображению с границами применялось выравнивание гистограммы, несколько морфологических операций, сглаживание медианным фильтром и преобразование водораздела как ключевой метод сегментации. Перебирая различные комбинации восстановленного изображения и его границ (вычисленным по одному и тому же изображению), можно добиться лучшего качества выявления границ. На изображениях на рис. 4 показаны различные комбинации восстановленного изображения и изображения с границами. Наилучшее качество выделения границ омматидиев демонстрирует изображение, представленное на рис. 4г, восстановленное по 500–6000 моделям с

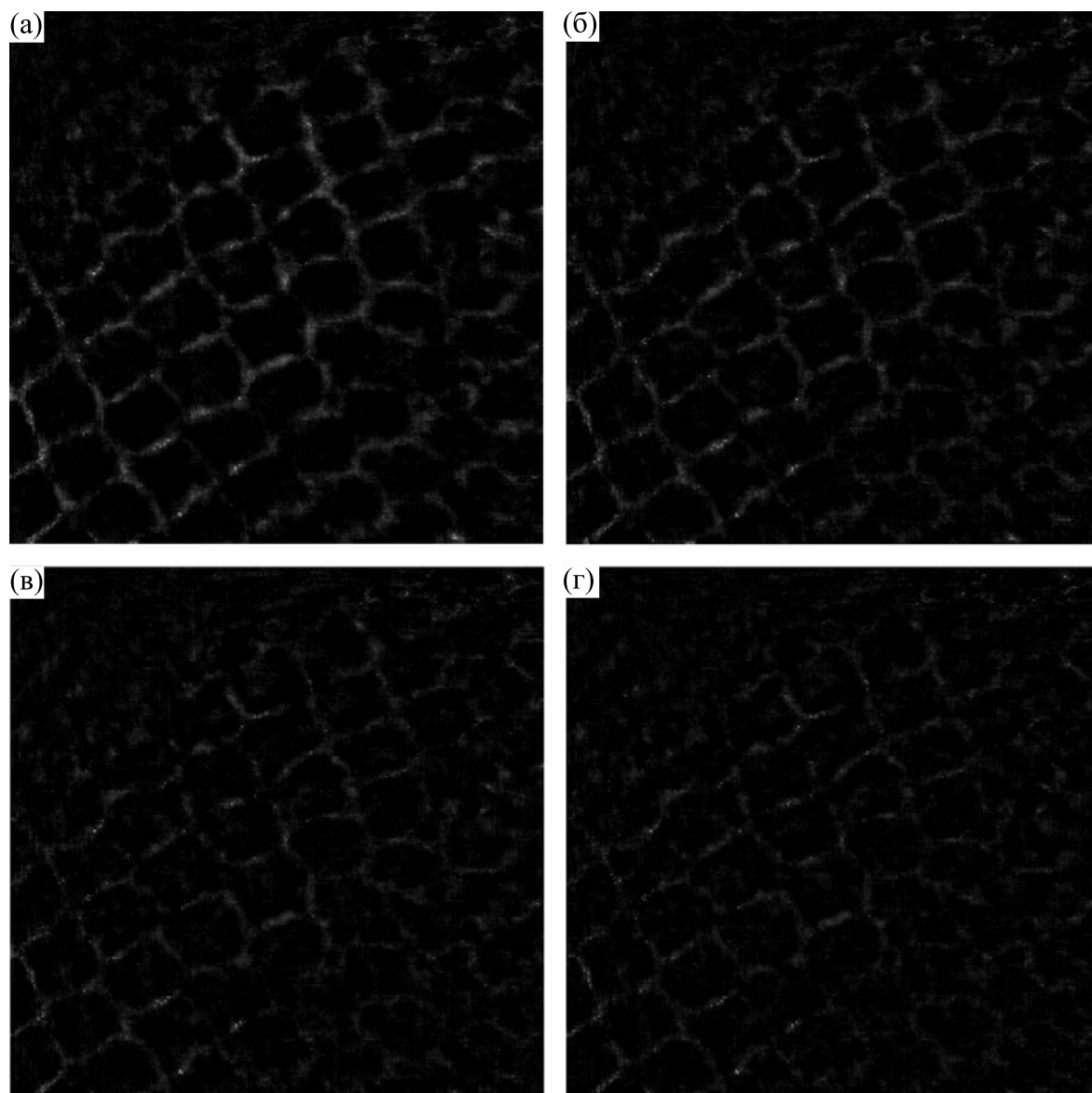


Рис. 2. Изображения, получаемые в результате исключения ранних частичных моделей: (а) — при исключении первых 500 моделей, (б) — при исключении первых 1000 моделей, (в) — при исключении первых 1500 моделей, (г) — при исключении первых 2000 моделей.

модулем градиента по 500–6000 моделям. Черные области по контуру изображения образуются из-за частичного обрезания объектов на краях, чего не происходит при анализе изображений всей области дифференцировки глазного диска.

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ МЕТОДА

Метод был реализован на языке программирования Си с использованием свободно доступной программы алгоритма ДГП с открытым исход-

ным кодом (<https://www.cgplibrary.co.uk/files2/About-txt.html>). Были реализованы методы чтения и предобработки исходных данных, возможность распараллеливания циклов, а также методы, реализующие непосредственно восстановление изображений (<https://github.com/Dannikk/Genetic-programming-feature-learning>). Для поиска границ была использована ранее разработанная авторами программа ProStack [18] (<https://sourceforge.net/projects/prostack/>).

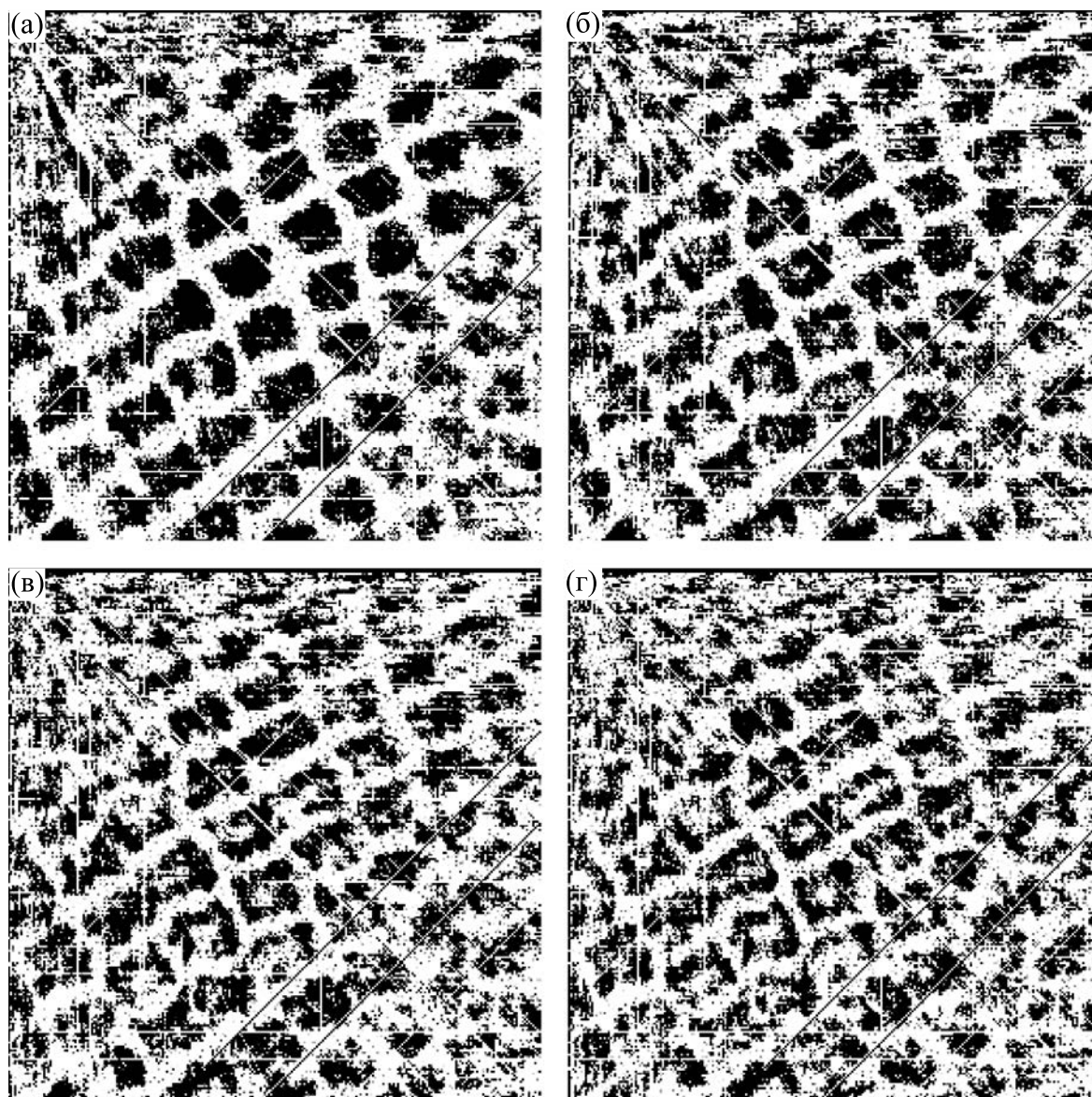


Рис. 3. Границы на изображении, получаемые в результате исключения ранних частичных моделей: (а) — при исключении первых 250 моделей; (б) — при исключении первых 500 моделей; (в) — при исключении первых 750 моделей; (г) — при исключении первых 1000 моделей.

ЗАКЛЮЧЕНИЕ

В данной работе разработан, реализован и применен к анализу конфокальных изображений метод, использующий парадигму ДГП. Метод успешно решает задачу выделения таких признаков, как границы омматидиев, на изображениях развивающегося глаза дрозофилы, окрашенных маркером на Fasciclin III. Улучшение качества обработки достигается путем отбрасывания первых частичных моделей в процессе восстановления изображений. Это позволяет добиться одновре-

менно сглаживания внутренних областей омматидиев и повышения контраста их границ.

ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 20-04-01047-а).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

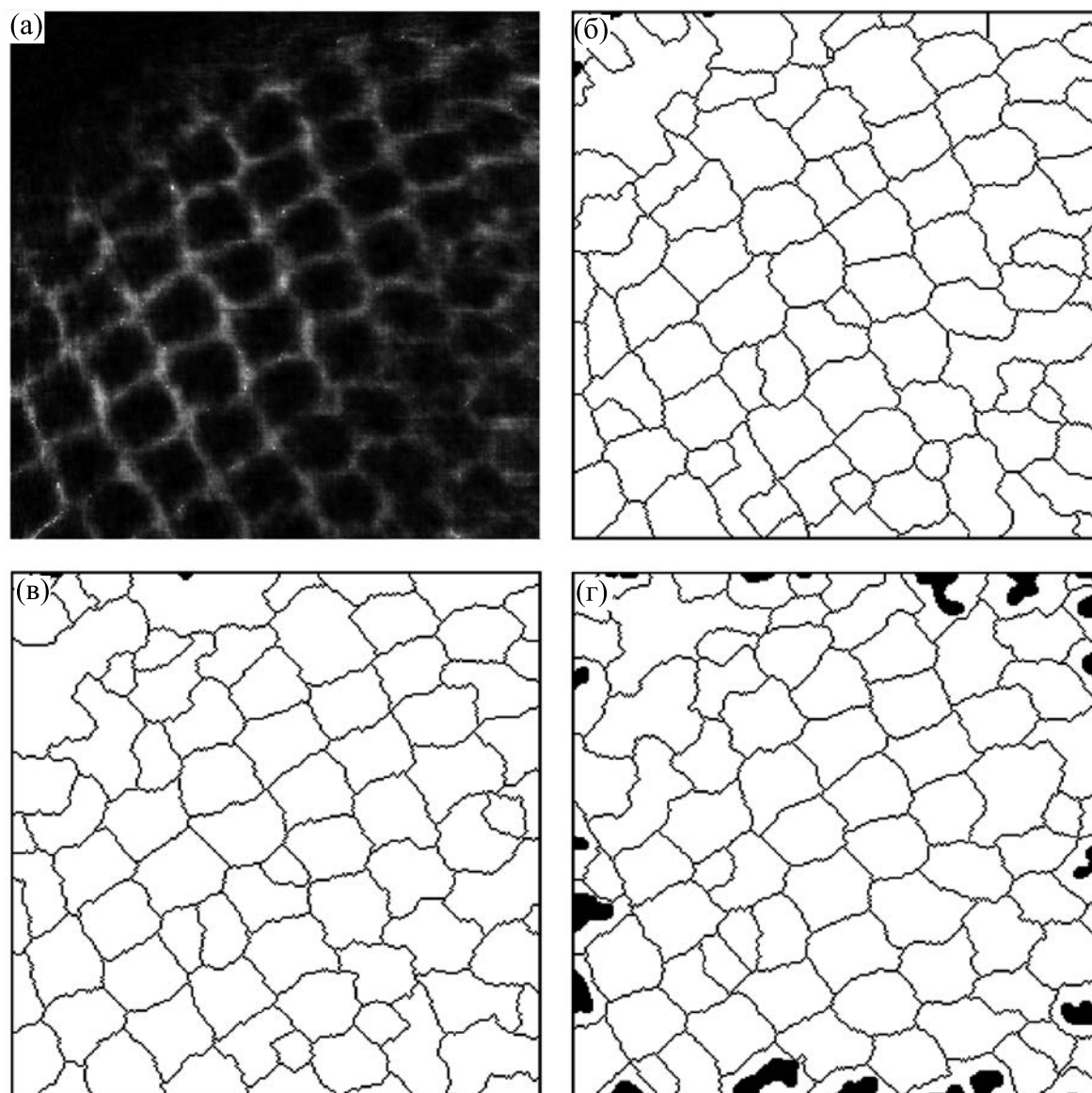


Рис. 4. Границы на изображении, получаемые в результате исключения ранних частичных моделей: (а) — восстановленное изображение; (б) — изображение восстановлено по 6000 моделей, модуль градиента по 250–6000 моделям; (в) — изображение восстановлено по 6000 моделей, модуль градиента по 500–6000 моделям; (г) — изображение восстановлено по 500–6000 моделям, модуль градиента по 500–6000 моделям.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания каких-либо исследований с использованием людей и животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. И. А. Русанова, в сб. матер. Всероссийской школы-семинара (Саратов, 01 октября 2018 г.), под ред. Д. А. Усанова (Изд-во «Саратовский источник», Саратов, 2018), сс. 78–81.
2. К. Н. Козлов, Е. В. Голубкова, Л. А. Мамон и др., *Биофизика*, 67, 283 (2022). DOI: 10.31857/S0006302922020119
3. J. P. Kumar, *Devel. Dynamics*, **241**, 136 (2012). DOI: 10.1002/dvdy.23707
4. S. Surkova, J. Görne, S. Nuzhdin, et al., *Devel. Biol.*, **476**, 41 (2021). DOI: 10.1016/j.ydbio.2021.03.005.
5. J. Y. Roignant and J. E. Treisman, *Int. J. Devel. Biol.*, **53**, 795 (2009). DOI: 10.1387/ijdb.072483jr
6. J. E. Treisman, *Wiley Interdisc. Rev. Devel. Biol.*, **2**, 545 (2013). DOI: 10.1002/wdev.100

7. S. Ali, S. A. Signor, K. Kozlov, et al., *Evolution & Development*, **21**, 157 (2019). DOI: 10.1111/ede.12283
8. L. Liu, L. Shao and X. Li, *Inf. Sci.*, **316**, 567 (2015). DOI: 10.1016/j.ins.2014.06.030
9. A. Lensen, H. Al-Sahaf, M. Zhang, et al., in *EuroGP 2016. LNCS*, Ed. by M. I. Heywood, J. McDermott, M. Castelli et al. (Springer, Cham, 2016), v. 9594, pp. 51–67. DOI: 10.1007/978-3-319-30668-1_4
10. S. Ruberto, V. Terragni, and J. Moore, in *Parallel Problem Solving from Nature. Lecture Notes in Computer Science Image Feature Learning with Genetic Programming* (Springer, Cham, 2020), pp. 63–78. DOI: 10.1007/978-3-030-58115-2_5
11. C. B. Perez and G. Olague, *Intell. Data Anal.*, **17**, 561 (2013). DOI: 10.3233/IDA-130594
12. W. A. Albukhanajer and J. A. Briffa, *IEEE Trans. Cybern.*, **45**, 1757 (2015). DOI: 10.1109/TCYB.2014.2360074
13. J. F. Miller, P. Thomson, and T.C. Fogarty, in *Genetic Algorithms and Evolution Strategies in Engineering and Computer Science: Recent Advancements and Industrial Applications*, Ed. by D. Quagliarella, J. Periaux, C. Poloni, and G. Winter (Wiley, 1998), pp. 105–131.
14. M. A. Kramer, *AIChE J.* **37**, 233 (1991). DOI: 10.1002/aic.690370209
15. A. Makhzani and B. J. Frey, in *Advances in Neural Information Processing Systems*, Ed. by C. Cortes, N. Lawrence, D. Lee, et al. (2015), pp. 2791–2799
16. P. Vincent, H. Larochelle, Y. Bengio, et al., in *Proc. Int. Conf. on Machine Learning, ICML 2008* (2008), pp. 1096–1103. DOI: 10.1145/1390156.1390294
17. P. M. Snow, A. J. Bieber, and C. S. Goodman, *Cell*, **59**, 313 (1989). DOI: 10.1016/0092-8674(89)90293-6
18. K. Kozlov, A. Pisarev, J. Kaandorp, et al., in *Abstr. Book of the 9th Int. Conf. Syst. Biol.* (Goteborg, 2008), p. 191.

Cartesian Genetic Programming for Image Analysis of the Developing Drosophila Eye

N.A. Danilov*, K.N. Kozlov*, S.Y. Surkova*, and M.G. Samsonova*

*Peter the Great St. Petersburg Polytechnic University, Polytekhnicheskaya ul. 29, St. Petersburg, 195251 Russia

Automatic feature extraction methods have gained increasing attention in modern image processing. The confocal images of the single-layered epithelium of the developing Drosophila eye may form an excellent model system to develop methods for complex feature extraction. The aim of this work was to explore Cartesian genetic programming for determination of the boundaries of ommatidia, the light-sensitive units in the presumptive eye region. Application of Cartesian genetic programming for the analysis of Fasciclin III expression has shown good results. This opens interesting perspectives for further use of this technology in the automatic analysis of confocal images.

Keywords: image analysis, Cartesian genetic programming, eye disc, Drosophila