

Визуализация данных об экспрессии генов на поверхности коры и гиппокампа мозга мыши

Ольга Сеньюкова
Лаборатория Компьютерной Графики и Мультимедиа,
Факультет Вычислительной Математики и Кибернетики,
Московский Государственный Университет, Москва, Россия
olsen222@yandex.ru

Аннотация

Задача исследования экспрессии генов в мозге чрезвычайно важна для понимания развития, функционирования и заболеваний мозга. Одним из этапов анализа экспрессии гена является визуализация. В данной работе визуализируется экспрессия генов в церебральной коре и гиппокампе, которая связана с механизмами памяти.

Исходными данными для визуализации являются трехмерная полигональная модель поверхности мозга (гиппокампа), набор трехмерных точек экспрессии некоторого гена, распределенных по объему коры (гиппокампа), а также атрибуты точек активности – сила экспрессии. Трехмерная визуализация экспрессии гена не всегда дает наглядное представление, поэтому рассматривается альтернативный подход – двумерная визуализация на поверхности анатомических структур. Точки экспрессии гена из указанного слоя (задается глубиной относительно поверхности и шириной) проецируются на поверхность мозга (гиппокампа).

Как правило, поверхностная визуализация применяется для данных функциональной МРТ и ПЭТ. В данной статье предлагается адаптация этих методов для визуализации данных об экспрессии генов.

Также предложен новый ускоренный алгоритм центрального проецирования точек экспрессии гена на поверхность.

Ключевые слова: *Экспрессия гена, Визуализация мозга, Карта мозга, Центральная проекция, Ортогональная проекция.*

1. ВВЕДЕНИЕ

Экспрессией гена называется процесс синтеза белка за счет активности генов в составе ДНК, содержащейся в ядре клетки. В каждой клетке экспрессируется лишь часть генов, которые и определяют ее структуру и функции. В мозге, в отличие от других органов, экспрессируются более 60% всех генов генома. Поэтому исследование активности генов в мозге является чрезвычайно важной задачей для понимания механизмов его работы.

Для исследования экспрессии определенного гена проводится эксперимент над лабораторным животным (обычно лабораторная мышь), в результате которого в его мозге начинает экспрессироваться данный ген. По окончании эксперимента у животного извлекают мозг и замораживают. Главная цель – определить, где и насколько активно экспрессировался исследуемый ген.

Для анализа используются тонкие срезы мозга (порядка 20 микрон). На срезы наносится специфический для данного гена молекулярный зонд. Визуализация зонда выявляет те клетки, где экспрессировался рассматриваемый ген.

После того, как получены срезы, они приводятся к каноническому виду – сопоставляются со специальным атласом мозга мыши (например, [1]). Затем по ним восстанавливается трехмерная модель мозга и трехмерные точки экспрессии гена.

Как правило, используется воксельная визуализация экспрессии генов [2]. Но такой способ визуализации не всегда позволяет качественно оценить данные из-за того, что экспрессия гена – в большом количестве отдельных точек, распределенных по объему структуры (Рисунок 1).

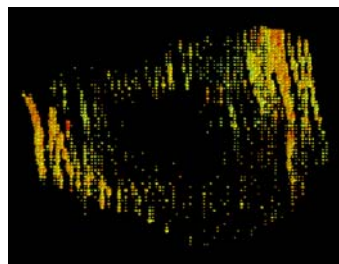


Рисунок 1: Объемная визуализация экспрессии гена.

Поэтому удобно использовать проекции данных экспрессии гена, находящиеся в указанном слое (задается глубиной относительно поверхности и шириной) на поверхность анатомических структур. Рассматривается два метода проецирования – центральное и ортогональное.

Результатом работы должна являться трехмерная модель поверхности мозга (гиппокампа) с нанесенной на нее текстурой – картой активности гена. Текстура поверхности генерируется таким образом, чтобы она была сглаженная, несмотря на дискретность исходных данных, и наглядно отображала картину экспрессии исследуемого гена.

Статья имеет следующую структуру. Часть 2 представляет собой краткий обзор существующих методов визуализации различных экспериментальных данных на поверхности мозга. В части 3 описаны алгоритмы проецирования точек экспрессии гена на поверхность заданной структуры (коры или гиппокампа). Часть 4 посвящена методам раскраски поверхности модели структуры по спроецированным на нее данным. Часть 5 посвящена описанию реализации и тестирования. Приведены примеры результатов визуализации. В заключительной 6 части указаны направления дальнейших исследований.

2. ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ

В [3] описано построение плоских карт для данных автордиографии мозга крысы. Для каждого среза строится разбиение на сектора с помощью лучей, идущих от центра среза. Для каждого сектора в каждом срезе рассчитывается среднее значение данных. Эти значения последовательно наносятся на двумерную карту. Если каждое значение рассматривать как цвет, то полученную карту можно использовать как текстуру модели мозга.

Авторы [4] предлагают методы поверхностной визуализации данных функциональной МРТ для мозга человека. Объемные экспериментальные данные (массив вокселей со значениями-цветами) визуализируются на полигональной модели поверхности мозга. Каждой вершине присваивается цвет по одному из двух методов. Цвета граней интерполируются по вершинам. Первый метод предполагает нахождение для каждой вершины поверхности мозга вокселя, в который она попадает, и присвоение вершине цвета этого вокселя. При отсутствии для данной вершины соответствующего вокселя используется трилинейная интерполяция ближайших вокселей для определения цвета вершины. Таким образом, этот метод позволяет визуализировать только те данные, которые находятся на поверхности мозга. Второй метод заключается в построении луча, идущего из центра мозга к заданной вершине, и присвоении вершине цвета, равного среднему арифметическому цветов вокселей, через которые проходит этот луч.

В [5] рассматривается проецирование различных видов экспериментальных данных (таких, как функциональная МРТ, позитронно-эмиссионная томография) на поверхность мозга человека. Данные представляются дискретно – в виде набора точек, распределенных по объему. В таком же виде (каждая точка – отдельным кружком) они и визуализируются на поверхности после проецирования. Метод проецирования заключается в поиске для каждой точки ближайшей вершины поверхности – по сути, приближенное ортогональное проецирование.

Таким образом, схема раскраски текстуры при поверхностной визуализации, как правило, соответствует виду начальных данных. Если данные непрерывные – в виде массива вокселей, то и текстура строится сглаженная. Если данные дискретные – в виде набора точек, то они и визуализируются на поверхности по отдельности.

В данной статье предлагается метод, который позволяет генерировать сглаженную текстуру поверхности по дискретным данным экспрессии гена.

3. АЛГОРИТМЫ ПРОЕЦИРОВАНИЯ ДАННЫХ НА ПОВЕРХНОСТЬ

Для построения текстуры точки экспрессии гена сначала проецируются на поверхность одним из двух методов – центральным (для мозга) или ортогональным (для мозга или гиппокампа). Попутно предложен новый метод центрального проецирования, который позволяет значительно ускорить процесс.

3.1 Центральное проецирование

При центральном проецировании данных на поверхность структуры в качестве центра проекции выбирается точка,

расположенная в центре исследуемой структуры. Проекцией точки экспрессии гена на поверхность считается пересечение луча, идущего из центра проекции и проходящего через данную точку, с поверхностью (Рисунок 2). Принадлежность точки указанному слою для проецирования будет определяться ее расстоянием до поверхности вдоль этого луча.

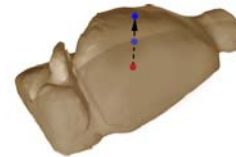


Рисунок 2: Схема центрального проецирования.

Для поиска проекции точки экспрессии гена на поверхность мозга проводится перебор не всех треугольников модели, а только треугольников из некоторой области. Это дает значительный прирост по скорости.

Область поиска нужного треугольника определяется следующим образом. Вводится эллиптическая система координат, начало которой помещается в центр мозга – центр проекции. Таким образом, каждую вершину поверхности мозга и каждую точку экспрессии гена можно задать в этой системе координат тройкой (λ, φ, ρ) , где λ и φ – углы, а ρ – расстояние до начала координат. Параметр ρ в данном алгоритме не используется, поэтому каждая точка задается только парой углов (λ, φ) (см. Рисунок 3)

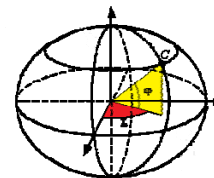


Рисунок 3: Эллиптическая система координат.

Пространство углов разбивается на ячейки – как показано на Рисунок 4.

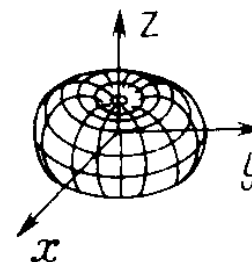


Рисунок 4: Разбиение пространства углов на ячейки.

Искомую область поиска составляют треугольники, вершины которых попадают в ту же ячейку, что и проецируемая точка экспрессии гена. Если в данной ячейке не найден треугольник, на который попадает проекция, то

рассматриваются ячейки, соседние с данной, затем – соседние с ними и т.д., пока не будет найдена проекция.

3.2 Ортогональное проецирование

При ортогональном проецировании для каждой точки экспрессии гена ищется ближайшая точка на поверхности (Рисунок 5).

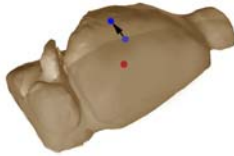


Рисунок 5: Схема ортогонального проецирования.

В данной работе используется приближенная схема ортогонального проецирования, аналогично [5], где для каждой точки исходных данных ищется не ближайшая точка, а ближайшая *вершина* поверхности. Принадлежность точки указанному слою для проецирования будет определяться ее расстоянием до этой вершины.

4. ВИЗУАЛИЗАЦИЯ ТРЕХМЕРНОЙ МОДЕЛИ

4.1 Визуализация экспрессии гена по точкам

Один из способов визуализации данных об экспрессии гена, спроецированных на поверхность мозга (гиппокампа) заключается в том, что каждая точка экспрессии гена изображена отдельным кружком, цвет которого зависит от силы экспрессии гена в данной точке (Рисунок 6(а-б)). Метод визуализации напоминает [5], но в [5] цвет не используется для кодирования значения данных в этой точке.

4.2 Сглаженная визуализация экспрессии гена

Предлагаемый метод сглаженной визуализации данных об экспрессии гена на поверхности мозга (гиппокампа) позволяет построить по спроецированным отдельным точкам сглаженную текстуру – картину экспрессии.

При построении сглаженной текстуры каждой вершине определенным образом присваивается цвет, чтобы затем цвета граней интерполировать по вершинам (как в [4]). Алгоритм присвоения цвета вершине зависит от метода, которым они были спроецированы точки экспрессии гена.

Если точки были спроецированы методом центрального проецирования, то каждой вершине присваивается цвет, соответствующий суммарной экспрессии гена в спроецированных точках, для которых эта вершина – ближайшая.

При приближенном ортогональном проецировании несколько точек могут проецироваться в одну и ту же вершину. Сила их экспрессии суммируется.

Таким образом, текстура поверхности получается сглаженная (Рисунок 6(б),7).

Данный способ визуализации позволяет наглядно выявлять области повышенной экспрессии гена, в то время как способ, описанный в п. 4.1 больше ориентирован на выявление местоположения всех точек экспрессии.

5. РЕАЛИЗАЦИЯ И ТЕСТИРОВАНИЕ

Алгоритмы реализованы в среде MATLAB.

5.1 Замеры производительности

Проведены замеры производительности предложенного алгоритма центрального проецирования. Для модели из 11948 треугольников и для 6304 точек экспрессии гена алгоритм работает 17 секунд. Для сравнения, алгоритм центрального проецирования с полным перебором всех треугольников модели работает около 30 минут. Таким образом, достигается значительный прирост производительности. Это весьма существенно для разработки интерактивного приложения, в котором пользователь может менять различные параметры и необходимо проводить проецирование заново.

Ускоренный алгоритм поиска пересечения луча с полигональной сеткой [7] работает 1 минуту 15 секунд, т.е. медленнее предложенного алгоритма более чем в 4 раза.

5.2 Результаты

В рамках данной работе визуализировались гены, экспрессирующиеся в церебральной коре мозга и гиппокампе – структуре в мозге, отвечающей за формирование памяти.

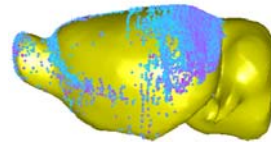


Рисунок 6(а): Визуализация экспрессии гена C-Fos в коре по точкам (метод проецирования - центральное).

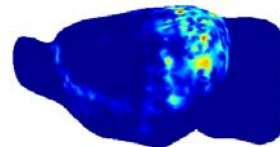


Рисунок 6(б): Сглаженная экспрессии гена C-Fos в коре по точкам (метод проецирования – приближенное ортогональное).

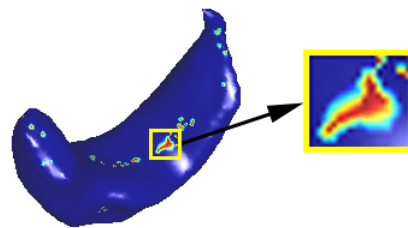


Рисунок 7: Сглаженная визуализация экспрессии гена Ctf1 в гиппокампе.

5.3 Интерактивное приложение

В среде MATLAB также разработано интерактивное приложение для визуализации данных об экспрессии генов на поверхности мозга и гиппокампа. Разработанный прототип обладает следующими возможностями:

- отображение трехмерной модели со спроецированными на нее точками активности гена (модель можно интерактивно вращать)
- выбор рассматриваемой структуры (кора или гиппокамп)
- выбор режима раскраски точек активности гена в зависимости от расстояния до поверхности или силы экспрессии
- задание процента отображаемых точек активности гена
- задание слоя для проецирования с помощью двух параметров – глубины (расстояния до поверхности) и ширины
- отображение исходных точек экспрессии гена (показать или скрыть).

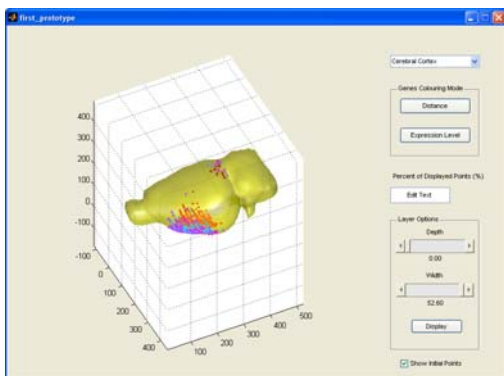


Рисунок 8: Прототип приложения

6. ДАЛЬНЕЙШИЕ ПЛАНЫ

В качестве ближайших планов по данной работе планируется построение плоской развертки структуры (мозга или гиппокампа) со спроецированными на нее данными – карты активности гена. Также планируется разработка методов анализа и сравнения данных. Помимо ортогонального и центрального проецирования будет рассматриваться метод проецирования вдоль колонок в мозге, которые будут распознаваться по данным оптической томографии.

7. БЛАГОДАРНОСТИ

Автор выражает благодарность Антону Конушину и Дмитрию Ветрову за помощь, оказанную во время работы. Автор также благодарит Константина Анохина за ценные консультации.

8. БИБЛИОГРАФИЯ

- [1] Allen Reference Atlas. ARA Coronal Atlas Images <http://mouse.brain-map.org/atlas/ARA/Coronal/browser.html>
- [2] C. Lau, L. Ng, C. Thompson, S. Pathak, L. Kuan, A. Jones, M. Hawrylycz. *Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain*. BMC Bioinformatics, 2008
- [3] D.P. Holschneidera, O.U. Scremine, D.R. Chialvog, B.P. Kayd, J.-M. I. Maarekd. *Flattened cortical maps of cerebral function in the rat: A region-of-interest approach to data sampling, analysis and display*. Neuroscience Letters, Vol. 434, Issue 2, pp.179-184, 2008

[4] P. Lincoln. *Surface projection method for visualizing volumetric data*. 2006 <http://sigpubs.biostr.washington.edu/archive/00000193>

[5] D. C. Van Essen, H. A. Drury, S. Joshi, M. I. Miller. *Functional and structural mapping of human cerebral cortex - Solutions are in the surfaces*. Proceedings of the National Academy of Sciences of the United States of America, Vol. 95, Issue 3, pp. 788-795, 1998

[6] H. Samet. *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, Reading, MA, 1990. ISBN 0-201-50255-0.

[7] K. Eaton. "boundary" class v2.1: a wrapper for surface objects

<http://www.mathworks.com/matlabcentral/fileexchange/20637>

Об авторе

Ольга Сенюкова – аспирант лаборатории Компьютерной Графики и Мультимедиа факультета Вычислительной Математики и Кибернетики Московского Государственного Университета им. М.В. Ломоносова. Ее электронный адрес - olsen222@yandex.ru.

Visualization of gene expression data on the surface of mouse brain cerebral cortex and hippocampus

Abstract

Exploring of gene expression in brain is very important for understanding of brain development, disease, and function. One of the stages of gene expression analysis is visualization. In this work genes expression in cerebral cortex and hippocampus which is essential for memory function, are visualized.

3D polygonal model of brain (hippocampus) surface, a set of expression points of a particular gene and points' attributes (expression scores) is initial data for visualization. 3D visualization is not always convenient. An alternative method – 2D visualization on the surfaces of anatomical structures – is regarded here. Gene expression points from a user-defined layer (which is set by depth – distance from the surface, and width) are projected onto the surface of the brain (hippocampus).

Usually surface-based visualization is used for fMRI and positron emission tomography. In this paper there is proposed an adaptation of these methods to visualization of gene expression data.

There is also proposed a novel fast algorithm of central projection.

Key words: Gene expression, Brain visualization, Surface-based visualization, Brain map, Central projection, Orthogonal projection.

About the author

Olga Senyukova is a PhD student at Moscow State University, Department of Computational Mathematics and Cybernetics, Graphics and Media Lab. Her email is olsen222@yandex.ru.