Computer Vision GraphiCon 2022

#### Биологически-подобная нейрона модель определения движения на накоплении сигнала

А.В. Кугаевских  $^{I}$ , А.Д. Мангараков  $^{I}$ 

### Аннотация

В статье представлена новая модель МТ-нейрона (нейрона средней височной области), которая позволяет обнаруживать движение и определять его направление и скорость без использования рекуррентной связи. Модель основана на накоплении сигнала и организована с использованием пространственно-временного вектора, который задает весовые коэффициенты. Пространственно-временной вектор формируется с помощью произведения гауссиана, задающего пространственную компоненту, и вейвлета «мексиканская шляпа», который задает временной вектор изменения рецептивного поля. Такая конфигурация позволяет не только детектировать движения, но и делать модель не чувствительной к равномерной или текстурной заливке. Модель представлена в вариантах для определения линейного и вращательного движения. Движение, в данном случае, представляет собой последовательную активацию нескольких нейронов выделения границ, расположенных в одном и том же направлении в определенной окрестности с течением времени, т.е. со сменой кадра. Для оценки движения модели проверялись на наборе данных MPI Sintel. Разработанная нами модель показывает результаты лучше, чем Spatio-Temporal Gabor. Наилучшая точность определения направления движения может быть получена при размере пространственно-временного вектора (7\*7,7).

### Ключевые слова

Определение движения, МТ-нейрон, биологически-подобная модель, нейронная сеть.

# Bio-inspired Neuron Model for Motion Detection on Base Signal Accumulation

A. V. Kugaevskikh <sup>1</sup>, A. D. Mangarakov <sup>1</sup>

ITMO University, Kronverksky Pr. 49., Saint-Petersburg, Russia

#### **Abstract**

The article presents a new model of the MT neuron (neuron of the middle temporal region), which allows motion detecting and determining its direction and speed without the use of recurrent communication. The model is based on signal accumulation and is organized using a space-time vector that sets the weighting coefficients. The space-time vector is formed using the product of the Gaussian, which defines the spatial component, and the "Mexican hat" wavelet, which sets the time vector of the change in the receptive field. This configuration allows not only to motion detect, but also to make the model not sensitive to uniform or textural fill. The model is presented in variants for determining linear and rotational motion. Motion, in this case, is the sequential activation of several edge selection neurons located in the same direction in a certain neighborhood over time i.e. with a change of frame. To assess the motion, the models were tested on the MPI Sintel dataset. The model developed by us shows results better than Spatio-Temporal Gabor. The best accuracy of determining the direction of movement can be obtained with the size of the space-time vector (7\*7, 7).

EMAIL: a-kugaevskikh@yandex.ru (A.B. Кугаевских), admangarakov@gmail.com (А.Д. Мангараков)

ORCID: 0000-0002-6676-0518 (А.В. Кугаевских), 0000-0002-8938-2214 (А.Д. Мангараков)

© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

19–22 September 2022, Ryazan, Russia

<sup>&</sup>lt;sup>1</sup> Университет ИТМО, Кронверкский пр. 49, Санкт-Петербург, Россия

ГрафиКон 2022: 32-я Международная конференция по компьютерной графике и машинному зрению, 19-22 сентября 2022 г., Рязанский государственный радиотехнический университет им. В.Ф. Уткина, Рязань, Россия

### **Keywords**

Motion detection, MT-neuron, bio-inspired model, neural network

## 1. Введение

В компьютерном зрении проблема анализа движения чаще всего решается с помощью применения уравнения оптического потока. При обучении нейронных сетей обнаружению движения мы также можем говорить об использовании уравнения оптического потока, или, скорее, базового механизма для определения направления изменения яркости пикселей. Нейронные сети также широко применяются в этой задаче. Чаще всего для выделения движения используют рекуррентные сети, такие как, GRUConv [1, 2], LSTMConv [3], в основе которых лежит комбинация пространственных (определяют положение объекта в кадре) и временных фильтров (определяют траекторию движения объекта с течением времени) с долгосрочными и краткосрочными фильтрами памяти, организованная в 3 слоя. Также встречается решение проблемы с помощью сверточных нейронных сетей [4-6], чаще всего это ResNet модель, лежащая в основе FlowNet [7] и FlowNet 2.0 [8]. Сети FlowNet выделяют вертикальные и горизонтальные составляющие смещения пикселей границ объектов в локальной окрестности аналогично алгоритму Лукаса-Канаде. Интересной вариацией является STCNN [9-11], обрабатывающая параллельно пространственный и временной потоки. При этом временной поток проходит 4 трехмерных сверточных слоя.

Неоднократно предпринимались попытки смоделировать нейрон выделения движения (МТ-нейрон) из зрительной коры. Наиболее распространенной моделью является модель Хигера-Симончелли [12, 13], которая подробно проанализирована в [14]. Модель Хигера-Симончелли основана на использовании энергии Габора, которая объединяет реальную и мнимую части фильтра Габора. Использование мнимой части разумно при обработке сигналов, так как позволяет компенсировать избыточность на низких частотах, но не имеет смысла при обработке изображений при свертке с яркостью пикселей. Другой подход был предложен в [15], где авторы используют 5 типов визуальных полей: горизонтальную и вертикальную скорость, разницу, контраст, внимание. Для формирования полей использованы фильтры Гаусса и Габора. Другой биологически-подобной моделью является модель LGMD [16], состоящая из тормозных связей, мембранного потенциала и основанная на изменении яркости пикселей, т.е. по сути оптическом потоке. Эти идеи были развиты в [17]. При этом LGMD модель не чувствительна к направлению движения. Еще одной попыткой можно считать использование спайковых нейронов [18], рецептивные поля которых задаются также с помощью фильтров Габора.

В данной работе предлагается альтернативный подход, основанный на биологическом подобии. В зрительной коре анализ движения начинается в первичной зрительной коре. Хотя его основная функция заключается в выделении краев, сложные клетки реагируют на движение в определенном направлении в пределах своего рецептивного поля. Более глубокий анализ движения выполняется в областях V3 и V5 (МТ) мозга. В конце концов, формируется общая карта движений в поле зрения.

### 2. Выделение краев

В предлагаемой модели выделение движения осуществляется на третьем слое нейронной сети. Первые два слоя осуществляют выделение краев, рисунок 1. Для этого была разработана нейронная сеть [19], основанная на использовании фильтра Габора и гиперболического тангенса. Изображения, поступающие на вход, представлена L\* компонентой пространства СІЕ L\*a\*b\*. На первом слое выделяются линии определенной ориентации. Второй слой отвечает за выделение комбинаций линий, в том числе и углов. Каждый слой содержит нейроны 3 типов, отличающихся конфигурацией рецептивных полей. При этом связи между слоями организованы особым образом. Каждый нейрон второго слоя ( $U_{C2}$ ) соединен только с двумя нейронами первого слоя ( $U_{S1}$ ). Таким образом, нейроны второго слоя позволяют выделять линии и углы (в случае фильтра Габора) и четырехугольники (в случае гиперболического тангенса).

Computer Vision GraphiCon 2022

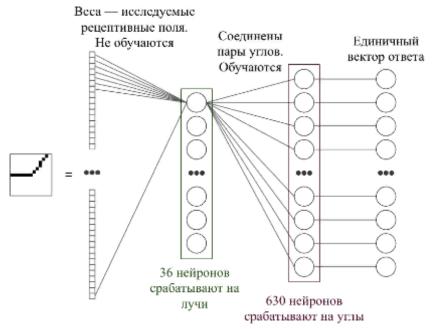


Рисунок 1 — Схема слоев выделения краев

Для выделения линий используются нейроны, рецептивное поле размера 7\*7 пикселей которых задается фильтром Габора.

которых задастех фильтром гаоора. 
$$G_{1,2} = exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \varphi\right), \tag{1}$$
 где  $\frac{\sigma}{\lambda} \approx 0,56$ ,  $\varphi$  – сдвиг фазы (0 – темная линия на светлом фоне,  $-\pi$  – светлая линия на темном

фоне),  $\gamma$  – степень эллиптичности (=0.1),  $\lambda$  = 3,  $\theta$   $\in$  [0,170].

Каждый нейрон чувствителен к линии определенной ориентации с шагом отклонения 10 градусов.

Для учета перепада яркости по области используется нейрон, рецептивное поле размера 17\*17 пикселей которого задается гиперболическим тангенсом.

$$G_3 = \tanh\left(\frac{2x'}{\lambda}\right),\tag{2}$$

На рисунке 2 представлены примеры рецептивных полей таких нейронов.

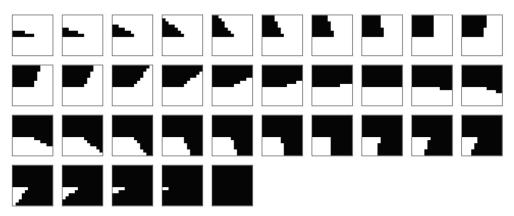


Рисунок 2 – Примеры выделяемых областей

Нейроны второго слоя представляют собой взвешенную сумму сигналов первого слоя с сигмоидальной функцией активации.

### 3. Модель МТ-нейрона

Обнаружение движение задается пространственно-временной схемой организации нейронной сети. Движение, в данном случае, представляет собой последовательную активацию нескольких нейронов выделения границ, расположенных в одном и том же направлении в определенной окрестности с течением времени, т.е. со сменой кадра. Таким образом, МТ-нейрон может определять направление  $\alpha$  и скорость  $\nu$  движения. МТ-нейрон, как и нейроны выделения краев, задается для каждого типа. Связи МТ-нейрона с нейронами  $U_{C2}$  соответствующего типа определяют рецептивное поле.

Для определения линейного движения рецептивное поле МТ-нейрона  $(U_{MT}^{\{l\}})$  включает последовательность  $U_{C2}$  нейронов в направлении  $\alpha$ . Для определения вращения рецептивное поле соответствующего МТ-нейрона  $(U_{MT}^{\{r\}})$  определяется связями с нейронами с совпадающим центром рецептивного поля, но отличающихся ориентацией  $\theta$ . Нейрон определения вращения создается дважды для разных направлений вращения.

$$U_{MT}^{\{l\}}(x, y, p, v, \alpha) = \sum_{x, y, t} U_{C2}(x, y, \theta, p) * w_{xy}(t),$$
(3)

$$U_{MT}^{\{r\}}(x, y, p, v, \alpha) = \sum_{\theta, t} U_{C2}(x_0, y_0, \theta, p) * w_{\theta}(t),$$
(4)

Веса МТ-нейронов задаются произведением гауссиана и вейвлета «мексиканская шляпа», первый отвечает за пространственную характеристику, второй задает коэффициент ослабления веса во времени.

$$w_{xy}(t) = exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \left(1 - \frac{t^2}{\sigma^2}\right) exp\left(-\frac{t^2}{2\sigma^2}\right),\tag{5}$$

$$w_{\theta}(t) = exp\left(-\frac{\theta^2}{2\sigma^2}\right)\left(1 - \frac{t^2}{\sigma^2}\right)exp\left(-\frac{t^2}{2\sigma^2}\right),\tag{6}$$

Равномерное заполнение такого нейрона неподвижной монотонной областью не приведет к активации нейрона. Коэффициент ослабления подчиняется определенному закону изменения t, рисунок 3: в начале движения в рецептивном поле t=[0,1,2], когда активируется нейрон  $U_{C2}(x_2,y_2,\alpha,p)$ , вектор t будет иметь значения t=[-1,0,1]. В конце движения вектор t будет иметь значения t=[-2,1,0]. В этом случае, значение коэффициента затухания меняется, рисунок 4.

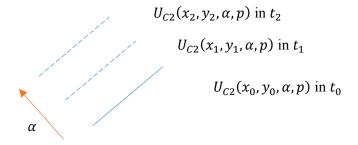
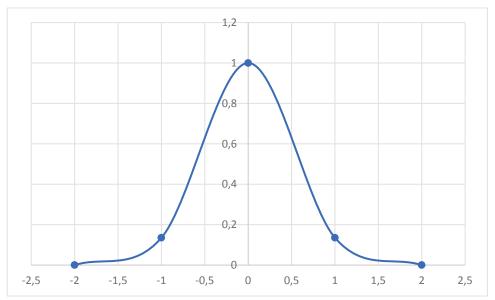


Рисунок 3 — Схема работы МТ-нейрона

Вейвлет «мексиканская шляпа» применяется для снижения активации при заполнении рецептивного поля МТ-нейрона текстурным периодическим изображением.

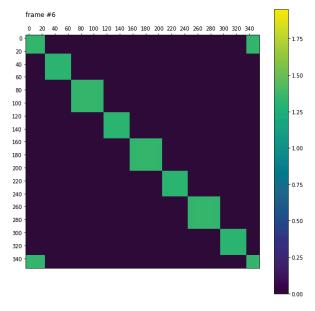
Computer Vision GraphiCon 2022



**Рисунок 4** – Изменение коэффициента затухания ( $\sigma = 0.5$ )

### 4. Экспериментальная проверка

Для экспериментальной проверки запустим движение в направлении 45 градусов. Рисунки. 5-7 показывают покадровую активацию МТ-нейронов. В идеальном случае должна быть тонкая линия, но из-за низкого разрешения появляется ложная активация в пределах 10-20 градусов ( $\sigma = 0.5$ ).



**Рисунок 5** — Первый кадр (начало движения, по вертикали — направление движения, по горизонтали — ответ МТ-нейрона)

Рисунок 5 показывает, что максимальная активация достигается на втором кадре, затем происходит затухание, что подтверждает наше предположение. Ослабление необходимо для того, чтобы МТ-нейрон не реагировал на неподвижные объекты.

В целом, чем дольше длится движение, тем точнее определяется его направление. Если мы увеличим размер рецептивного поля пространственно-временного вектора с 3 до 7, точность определения движения повышается, рисунок 8-9.

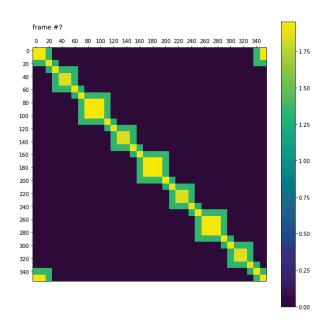


Рисунок 6 - Второй кадр

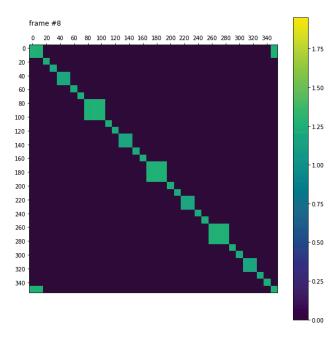


Рисунок 7 - Третий кадр (конец движения)

Тем или иным образом все модели анализа движения строят оптический поток, что позволяет провести их сравнение по метрикам AEE (средняя оконечная ошибка) и AAE (средняя угловая ошибка), результаты представлены в таблице 1. Для оценки движения модели проверялись на наборе данных MPI Sintel [20]. Разработанная нами модель не обеспечивает ярко выраженной реакции на изменение скорости, поэтому для нее не проводились измерения метрики AEE, что можно признать недостатком модели по сравнению с остальными.

$$AEE = \frac{1}{|\Omega|} \sum_{\Omega} |\vec{v}_{ref} - \vec{v}_{est}| \tag{7}$$

$$AAE = \frac{1}{|\Omega|} \sum_{\Omega} a\cos\left(\frac{\vec{v}_{ref} * \vec{v}_{est}}{|\vec{v}_{ref}| |\vec{v}_{est}|}\right), \tag{8}$$

где  $\Omega$  -множество векторов.

Computer Vision GraphiCon 2022

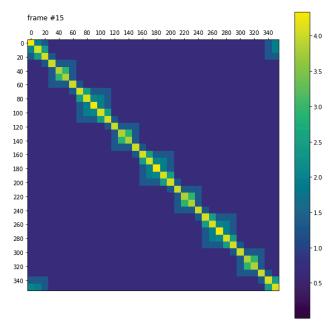


Рисунок 8 – Четвертый кадр

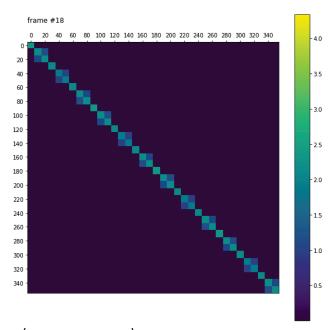


Рисунок 9 – Седьмой кадр (конец движения)

Таблица 1 – Сравнение качества выделения движения

		те на теотра рогдения долинения
Название модели	AAE	AEE
FlowNetS	0.467592	14.322449
FlowNetC	0.4421572	12.828806
Higer-Simoncelli	1.156201	41.9602013
ST-Gabor	1.609959	42.1128158
Optical Flow (Farneback)	0.520226	25.1728420
Our model	1.373866	Не чувствительна

### 5. Заключение

Представленная модель нейрона МТ не реагирует на неподвижный объект, так как равномерное заполнение рецептивного поля такого нейрона дает выходное значение, максимально близкое к нулю. Движение кодируется его направлением и скоростью.

Наилучшая точность определения направления движения может быть получена при размере пространственно-временного вектора (7\*7,7).

Эксперименты показали, что предложенная модель нейрона МТ реагирует на движение ожидаемым образом с точностью, схожей с другими биологически-подобными моделями по метрике определения ориентации. Несмотря на то, что обучаемые нейронные сети (LSTM, FlowNet) дают более качественное обнаружение движения, их недостаток заключается как раз в обучаемости, что снижает универсальность и не позволяет использовать эти модели на самих камерах в системах распределенного компьютерного зрения. При высокой разрешающей способности камер точности биологически-подобных моделей достаточно для определения направления и скорости движения. Разработанная нами модель МТ-нейрона уступает по качеству аналогичной модели Хигера-Симончелли.

### 6. Список источников

- [1] Cai Y. et al. Video anomaly detection with multi-scale feature and temporal information fusion // Neurocomputing. 2021. Vol. 423. P. 264–273. DOI: 10.1016/j.neucom.2020.10.044.
- [2] Tokmakov P., Schmid C., Alahari K. Learning to Segment Moving Objects // Int J Comput Vis. 2019. Vol. 127, № 3. P. 282–301. DOI: 10.1007/s11263-018-1122-2.
- [3] Szeto R. et al. A Temporally-Aware Interpolation Network for Video Frame Inpainting // IEEE Trans. Pattern Anal. Mach. Intell. 2020. Vol. 42, № 5. P. 1053–1068. DOI: 10.1109/TPAMI.2019.2951667.
- [4] Kim D. et al. Recurrent Temporal Aggregation Framework for Deep Video Inpainting // IEEE Trans. Pattern Anal. Mach. Intell. 2020. Vol. 42, № 5. P. 1038–1052. DOI: 10.1109/TPAMI.2019.2958083.
- [5] Wang H. et al. Skeleton edge motion networks for human action recognition // Neurocomputing. 2021. Vol. 423. P. 1–12. DOI: 10.1016/j.neucom.2020.10.037.
- [6] Maczyta L., Bouthemy P., Le Meur O. CNN-based temporal detection of motion saliency in videos // Pattern Recognition Letters. 2019. Vol. 128. P. 298–305. DOI: 10.1016/j.patrec.2019.09.016.
- [7] Dosovitskiy A. et al. FlowNet: Learning Optical Flow with Convolutional Networks // 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015. P. 2758–2766. DOI: 10.1109/ICCV.2015.316.
- [8] Ilg E. et al. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017. P. 1647–1655. DOI: 10.1109/CVPR.2017.179.
- [9] Jing C. et al. Spatiotemporal neural networks for action recognition based on joint loss // Neural Comput & Applic. 2020. Vol. 32, № 9. P. 4293–4302. DOI: 10.1007/s00521-019-04615-w
- [10] Chen G. et al. Spatial-Temporal Attention-Aware Learning for Video-Based Person Re-Identification // IEEE Trans. on Image Process. 2019. Vol. 28, № 9. P. 4192–4205. DOI: 10.1109/TIP.2019.2908062.
- [11] Miao Y. et al. ST-CNN: Spatial-Temporal Convolutional Neural Network for crowd counting in videos // Pattern Recognition Letters. 2019. Vol. 125. P. 113–118. DOI: 10.1016/j.patrec.2019.04.012.
- [12] Heeger D.J. Model for the extraction of image flow // J. Opt. Soc. Am. A. 1987. Vol. 4, № 8. P. 1455. DOI: 10.1364/josaa.4.001455.
- [13] Simoncelli E.P., Heeger D.J. A model of neuronal responses in visual area MT // Vision Research. 1998. Vol. 38, № 5. P. 743–761. DOI: 10.1016/S0042-6989(97)00183-1.
- [14] Chessa M., Sabatini S.P., Solari F. A systematic analysis of a V1–MT neural model for motion estimation // Neurocomputing. 2016. Vol. 173. P. 1811–1823. DOI: 10.1016/j.neucom.2015.08.091.

Computer Vision GraphiCon 2022

[15] Rezai O., Boyraz Jentsch P., Tripp B. A video-driven model of response statistics in the primate middle temporal area // Neural Networks. 2018. Vol. 108. P. 424–444. DOI: 10.1016/j.neunet.2018.09.004.

- [16] Yue S. et al. Reactive direction control for a mobile robot: a locust-like control of escape direction emerges when a bilateral pair of model locust visual neurons are integrated // Auton Robot. 2010. Vol. 28, № 2. P. 151–167. DOI: 10.1007/s10514-009-9157-4.
- [17] Xu J., Park S.H., Zhang X. A bio-inspired motion sensitive model and its application to estimating human gaze positions under classified driving conditions // Neurocomputing. 2019. Vol. 345. P. 23–35. DOI: 10.1016/j.neucom.2018.09.093.
- [18] Tsur E.E., Rivlin-Etzion M. Neuromorphic implementation of motion detection using oscillation interference // Neurocomputing. 2020. Vol. 374. P. 54–63. DOI: 10.1016/j.neucom.2019.09.072.
- [19] Kugaevskikh A.V., Sogreshilin A.A. Analyzing the Efficiency of Segment Boundary Detection Using Neural Networks // Optoelectron.Instrument.Proc. 2019. Vol. 55, № 4. P. 414–422. DOI: 10.3103/S8756699019040137.
- [20] MPI Sintel Flow Dataset [Электронный ресурс]. URL: http://sintel.is.tue.mpg.de/. (дата обращения 12.08.2022).