

РАЗРАБОТКА МЕТОДА ПОИСКА ОБЛАСТЕЙ ОТКРЫТИЯ/ЗАКРЫТИЯ НА ОСНОВЕ РЕКУРРЕНТНОЙ НЕЙРОСЕТЕВОЙ АРХИТЕКТУРЫ

Великанов Максим Сергеевич

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: maksim.velikanov@graphics.cs.msu.ru

Научный руководитель — Ватолин Дмитрий Сергеевич

При обработке видеопоследовательностей часто возникает задача сопоставления кадров. Многие методы обработки видео используют соответствия между кадрами для стабилизации работы по времени, например методы сегментации движущихся объектов, методы колоризации, методы оценки качества видео, методы матирования видео. Также сопоставление необходимо между ракурсами стереовидео для вычисления диспаратности. Трудности для алгоритмов обработки видео составляют так называемые области открытия/закрытия – пиксели, присутствующие в одном кадре и отсутствующие в другом, поэтому данные области не стоит учитывать. Известно, что задача оценки оптического потока (движения между кадрами) тесно связана с задачей построения областей открытия/закрытия.

В данной работе за основу взят нейросетевой метод вычисления оптического потока RAFT [1]. Частью архитектуры нейросети является кодировщик, извлекающий признаки из обоих кадров. Признаки скалярно перемножаются, составляется 4-х мерная карта сопоставления ошибок, которая затем уменьшается в 1, 2, 4 и 8 раз для представления соответствий между признаками на разных масштабах. Основой архитектуры является оператор обновления, реализующий итерационное приближение оптического потока. С помощью оператора обновления происходит сопоставление признаков одновременно на всех масштабах. При увеличении числа итераций результат сходится к итоговому оптическому потоку.

В работе предлагается изменить архитектуру RAFT для одновременной оценки оптического потока и областей открытия/закрытия следующим образом. На каждой итерации по карте признаков GRU-модуля вычисляется карта областей открытия/закрытия, складывается с предыдущим приближением областей и подается на вход оператору обновления в следующей итерации. Разрешение результата на каждой итерации составляет $1/8$ от изначального. Для восстановления изначального разрешения оптический поток и области

открытия/закрытия подаются в цепочку из 6 residual-блоков, после каждого второго из них разрешение увеличивается в 2 раза. В качестве дополнительной информации на вход 1, 3 и 5 блоку подается конкатенация признаков первого кадра и признаков второго кадра, интерполированных по оптическому потоку.

Функция ошибки для обучения нейросети учитывает оптический поток и области открытия/закрытия.

$$L = \sum_{i=1}^N \gamma^{N-i} (w_{flow} L_{flow} + L_{occ}), \quad (1)$$

где N – количество итераций, $\gamma = 0.8$, L_{flow} – функция ошибки оптического потока, L_{occ} – функция ошибки областей открытия/закрытия, w_{flow} – балансирующий коэффициент, приравнивающий значения слагаемых.

В функцию ошибки входит результат каждой итерации оператора обновления с увеличением веса последних итераций. Для областей открытия/закрытия используется взвешенная бинарная кросс-энтропия:

$$L_{occ} = - \sum_x (\alpha o_1(x) \log o_{1,gt}(x) + \beta (1 - o_1(x)) \log (1 - o_{1,gt}(x))) \quad (2)$$

$$\alpha = \frac{H \cdot W}{\sum_x o_1(x) + \sum_x o_{1,gt}(x)}, \beta = \frac{H \cdot W}{\sum_x (1 - o_1(x)) + \sum_x (1 - o_{1,gt}(x))}, \quad (3)$$

где x – пиксель, o_1 – результат вычисления в диапазоне $[0, 1]$, $o_{1,gt}$ – истинное значение, H – высота кадра, W – ширина кадра.

Обучение производилось в несколько этапов: на наборах данных FlyingChairsOcc [3] и FlyingThings3D [4] по 100 000 итераций на каждом. FlyingChairsOcc состоит из 22872 кадров разрешения 1024×768 . Набор данных FlyingThings3D составляет 21818 кадров разрешения 960×540 , в нем движения имеют уже более сложный характер. Оба набора данных содержат исходные кадры, истинные значения оптического потока и областей открытия/закрытия. Полученная модель тестировалась на наборе данных Sintel Final, качество областей открытия/закрытия оценивалось по F1-мере. F1-мера предложенного метода – 0.712, точность – 0.742, полнота – 0.684. Пример постро-

енных предложенным методом областей открытия/закрытия представлен на рис. 1.

Исследования выполнены при финансовой поддержке РФФИ в рамках научного проекта 19-01-00785 а.

Иллюстрации



Рис. 1. Сверху: кадр из тестового набора Sintel Final. Справа: истинные области открытия/закрытия. Слева: области, вычисленные предложенным методом

Литература

1. Teed Z., Deng J. Raft: Recurrent all-pairs field transforms for optical flow //European Conference on Computer Vision. – 2020. – С. 402-419
2. Sun D., Yang X., Liu M. Y., Kautz J. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – С. 8934–8943.
3. Hur J., Roth S. Iterative residual refinement for joint optical flow and occlusion estimation //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2019. – С. 5754-5763.
4. Mayer N., Ilg E., Hausser P., Fischer P., Cremers D., Dosovitskiy A., Brox T. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation //Proceedings of the IEEE Conference on Computer Cision and Pattern Recognition. – 2016. – С. 4040-4048.