

Нейросетевой алгоритм поиска областей открытия/закрытия в видеопоследовательностях

М. С. Великанов, А. Б. Анзина, С. В. Лаврушкин, Д. С. Ватолин

Аннотация—Области открытия/закрытия по движению представляют собой набор пикселей, которые видны только в одном из двух подряд идущих кадров. Поиск областей открытия/закрытия является важной задачей компьютерного зрения. Точное определение этих областей позволит улучшить качество многих методов обработки видео, таких как: интерполяция кадров, оценка оптического потока, распространение цвета и других. Большинство существующих методов основано на оптимизации функции энергии, что приводит к большим вычислительным затратам. Также стоит учитывать, что качественное нахождение областей открытия/закрытия затруднительно без какой-либо информации о движении между кадрами, но и при оценке оптического потока знание данных областей позволяет избежать ложных соответствий между пикселями кадров. Исходя из этих рассуждений в рамках данного исследования был разработан новый метод поиска областей открытия/закрытия, основанный на алгоритме построения карт оптического потока PWC-net. Ключевой идеей является построение пирамиды признаков с разными разрешениями для обработки кадров. Данный способ обработки заимствован из стандартных методов компьютерной графики и хорошо зарекомендовал себя. Также было проведено экспериментальное сравнение разработанного метода с 15 аналогами на наборе данных MPI-Sintel.

Ключевые слова—Глубокое обучение, нейросетевые методы, области открытия/закрытия, обработка видео.

I. Введение

Области открытия/закрытия играют важную роль в оценке диспаратета и оптического потока. Исключение данных областей из обработки улучшит существующие алгоритмы сопоставления кадров, так как это позволит избавиться от шума в областях открытия/закрытия. Кроме того, области открытия/закрытия важны для задачи сегментации объектов в кадре, обнаружения границ по глубине и границ по движению. Неправильное определение областей открытия/закрытия, особенно на границах объектов, может серьезно ухудшить качество во многих прикладных задачах оценки оптического потока, например, в интерполяции кадров [1], [2], в сегментации

Статья получена 27 ноября 2019. Работа основана на результате магистерской диссертации. Исследования выполнены при финансовой поддержке РФФИ в рамках научного проекта 19-01-00785 а. Обучение моделей производилось на вычислительном комплексе IBM Polus факультета вычислительной математики и кибернетики МГУ: <http://hpc.cmc.msu.ru/polus>.

Максим Сергеевич Великанов, МГУ им. М. В. Ломоносова (email: maksim.velikanov@graphics.cs.msu.ru).

Александра Борисовна Анзина, МГУ им. М. В. Ломоносова (email: alexandra.anzina@graphics.cs.msu.ru).

Сергей Валерьевич Лаврушкин, МГУ им. М. В. Ломоносова (email: sergey.lavrushkin@graphics.cs.msu.ru).

Дмитрий Сергеевич Ватолин, МГУ им. М. В. Ломоносова (email: dmitriy@graphics.cs.msu.ru).

движения [3], [4] и в кодировании с компенсацией движения [5], [6].

Области открытия/закрытия представляют собой набор пикселей, которые видны только в одном из двух подряд идущих кадров в видеопоследовательности или одном из двух ракурсов в стереовидео.

Стоит отметить, что оценка оптического потока и областей открытия/закрытия является хорошо известной проблемой курицы и яйца [7], [8], [9]. Точное определение рассматриваемых областей имеет решающее значение для надежной оценки оптического потока. Тем не менее рассматриваемые области являются следствием движения, и оценка точного оптического потока, наоборот, необходима для надежного определения этих областей. Таким образом, исходя из их взаимосвязи, следует, что для получения наилучшего результата требуется одновременно работать как с оптическом потоком, так и с областями открытия/закрытия.

В данной работе предложен новый нейросетевой метод поиска областей открытия/закрытия. В секции II описываются ранее разработанные методы поиска областей открытия/закрытия, а также методы поиска оптического потока и семантической сегментации изображений. Далее, в секции III приводится базовый метод, вычисляющий области открытия/закрытия с использованием оптического потока, и исследуется применимость методов сегментации изображений. Секции IV и V содержат детали предложенного метода и сравнение его с остальными рассмотренными методами.

II. Обзор существующих методов

Существует большое количество подходов к решению задачи построения областей открытия/закрытия, многие из которых тесно связаны с задачей оценки оптического потока.

A. Оптимизационные методы

В статье [10] предлагается симметричный метод оценки прямого и обратного оптического потока одновременно с поиском областей открытия/закрытия. Основная идея алгоритма сводится к оптимизации функции энергии, которая включает в себя фотометрическую ошибку, проверку траекторий движения суперпикселей, вычисление евклидова расстояния между положением одного и того же пикселя в соседних кадрах и проверку соответствия областей открытия/закрытия. В статье [11] также

решается задача оптимизации функции энергии, которая применяется к суперпикселям. Метод включает в себя вычисление функции реконструкции на основе карты соответствий суперпикселей, полученной нейросетевым методом DeepMatching [12].

Не все методы построения областей открытия/закрытия решают исключительно стандартную задачу оптимизации функции энергии. Например, в статье [13] описан метод с использованием марковской цепи для улучшения точности поиска областей открытия/закрытия. Кроме того, существуют алгоритм поиска областей открытия/закрытия, который использует ансамбль созданных вручную функций извлечения признаков [14], метод, основанный на анализе изменения цвета в зависимости от оптического потока [15], метод, основанный на разбиении изображения на слои и упорядочивания их по глубине [16] и метод, использующий информацию о взаимосвязи локальных слов [17].

Так как оптимизационные методы работают не менее 100 секунд на пару кадров, их использование на практике сильно ограничено. Поэтому все большую популярность набирают нейросетевые методы.

В. Нейросетевые методы

На момент написания статьи известно всего два нейросетевых метода построения областей открытия/закрытия.

Первый метод [18] представляет собой симметричную сверточную нейронную сеть SymmNet, разработанную для нахождения областей открытия/закрытия между левым и правым ракурсами в стереовидео. Эту архитектуру можно применить для нахождения областей открытия/закрытия по движению и в обычном 2D видео.

Второй метод [19] основывается на нейросетевой архитектуре FlowNet. Отличительной особенностью данного метода является то, что обучение происходит не только на областях открытия/закрытия, но и на диспаратете, оптическом потоке, границах по движению и границах по глубине.

Отдельно стоит упомянуть нейросетевые методы для семантической сегментации изображений. В частности, такие нейросетевые архитектуры, как U-Net [20], Link-net [21] и BiSeNet [22], можно применять для предсказания бинарной маски изображения, что позволяет их использовать для попиксельной классификации изображений на наличие или отсутствие областей открытия/закрытия.

С. Нейросетевые методы построения оптического потока

Наиболее популярным методом для оценки оптического потока на момент написания статьи является FlowNet 2.0 [23]. В основе разработанной авторами архитектуры лежат нейросетевые модели FlowNetS и FlowNetC [24], соединенные последовательно.

В методе [25] используется нейросеть, которая встраивает исходное изображение в компактное евклидово пространство для вычисления оптического потока. На выходе нейросети получается разреженный оптический поток, который необходимо интерполировать

локально взвешенной аффинной моделью EpicFlow [26]. Стоит отметить, что в этом методе для обработки областей открытия/закрытия используется интерполяция с учетом границ.

PWC-net, описанная в [27], появилась как продолжение иерархического подхода в обработке изображений и видео. В классических методах вычисление оптического потока проводится для нескольких разрешений исходного изображения, начиная с самого маленького. Вместо уменьшенных копий исходных изображений в PWC-net используются нейросетевые признаки, полученные с помощью сверток. В отличие от FlowNet 2.0, PWC-net имеет существенно меньше параметров, следовательно быстрее обучается и обладает большей производительностью.

III. Исследование применимости методов семантической сегментации

Задачу построения областей открытия/закрытия можно свести к попиксельной классификации для каждой пары подряд идущих кадров, следовательно, к этой задаче можно применить популярную архитектуру для семантической сегментации U-Net [20]. Также исследование проводилось для Link-net [21], где в качестве кодировщика и декодировщика была использована предобученная на ImageNet [28] архитектура ResNet18 [29], и для BiSeNet [22], являющейся на момент написания статьи лучшей архитектурой для сегментации изображений. Для U-Net и Link-net в качестве функции потерь во время обучения была выбрана бинарная кросс-энтропия, а для BiSeNet — коэффициент Серенсена.

Обучение всех исследуемых нейросетей производилось на наборах данных FlyingChairs2 [19] и FlyingThings3D [30]. В FlyingThings3D отсутствуют истинные значения областей открытия/закрытия, но их можно вычислить с помощью имеющихся в нем истинных карт сегментации и оптического потока. Для тестирования использовался набор данных MPI-Sintel (Final), в котором есть истинные значения областей открытия/закрытия.

Лучший результат был получен для U-Net с F-мерой 0.357 (Таблица I), что сильно отстает от результата, полученного с помощью адаптированного LRC. Таким образом, рассмотренным нейросетевым архитектурам не хватает информации об оптическом потоке для построения качественных областей открытия/закрытия.

Таблица I. Сравнение нейросетевых архитектур для семантической сегментации на наборе данных MPI-Sintel (Final).

Архитектура	U-Net	Link-net	BiSeNet
F-мера	0.357	0.147	0.124

IV. Предложенный метод

В качестве точки отсчета будем использовать простой алгоритм на основе оптического потока, полученного нейросетевым методом. Применив к картам прямого и обратного оптического потока адаптированный для движения LRC [31], получим искомые области открытия/закрытия. Так как каждый пиксель классифицируется либо как принадлежащий, либо как не принадлежащий областям открытия/закрытия,

для оценки результата воспользуемся F-мерой. Соответствующие результаты для нейросетей PWC-net и FlowNet 2.0 можно увидеть в Таблице II. Перейдем

Таблица II. Сравнение базовых алгоритмов.

Метод	Точность	Полнота	F-мера
PWC-net	0.470	0.489	0.479
FlowNet 2.0	0.509	0.467	0.487

к описанию предложенного метода. В качестве основы взята архитектура PWC-net, изначально разработанная для задачи построения оптического потока. Отличительной особенностью этой архитектуры является использование пирамиды извлечения признаков и последовательного построения карты оптического потока, начиная с низкого разрешения. Общая схема разработанного нейросетевого метода представлена на Рис. 1.

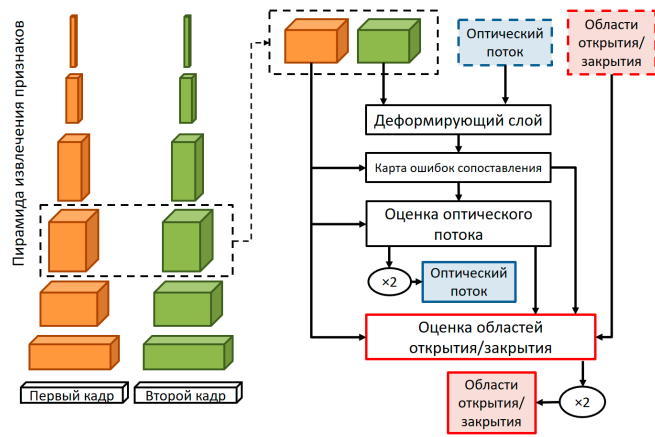


Рис. 1. Общая схема разработанного метода. Оптический поток и области открытия/закрытия, обведенные пунктирной линией, получены с предыдущего уровня пирамиды. Увеличение разрешения обозначено на схеме как $\times 2$.

Главная модификация архитектуры PWC-net состоит в том, что на каждом уровне пирамиды добавляется небольшая сверточная нейронная сеть, состоящая из 6 слоев с глубинами 128, 128, 96, 64, 32, 2 и размером ядра 3×3 . После каждого сверточного слоя используется функция активации Leaky ReLU [32] и конкатенация (Рис. 2) с результатом предыдущего блока, а на последнем уровне пирамиды добавляется Softmax-слой после увеличения разрешения при помощи свертки с транспонированным ядром. Используемая цепочка конкатенаций аналогична Dense-блоку из архитектуры DenseNet [33].

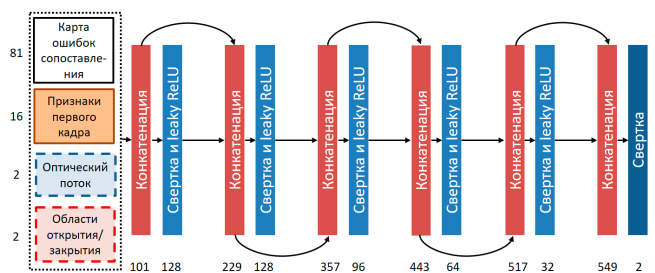


Рис. 2. Архитектура сверточной нейросети для оценки областей открытия/закрытия. Указанное количество каналов может меняться в зависимости от уровня пирамиды.

Обучение предложенного нейросетевого метода осуществлялось с использованием обучающих данных, описанных в секции III, где из каждой пары кадров случайным образом вырезался небольшой фрагмент с разрешением 256×256 пикселей. Веса, отвечающие за оценку оптического потока, кроме весов с 6-го уровня пирамиды, были проинициализированы значениями из предобученной модели PWC-net и во время оптимизации замораживались с целью уменьшения времени обучения. В статье [27] хоть и описаны 6 уровней пирамиды, авторы использовали только последние 5 уровней, из-за чего выходное разрешение падает в 16 раз относительно входного. Предложенная архитектура включает в себя все 6 уровней, поэтому разрешение уменьшается только в 4 раза, но это изменение повлекло увеличение общего числа параметров нейросети.

В качестве оптимизируемой функции во время обучения использовалась бинарная кросс-энтропия на каждом уровне пирамиды нейронной сети:

$$L = \sum_{p=1}^P \alpha_p \left(\sum_x I[O(x) = 1] \log(\sigma(x)) + \sum_x I[O(x) = 0] \log(1 - \sigma(x)) \right),$$

где x — пиксель изображения, $O(\cdot)$ — проверка принадлежности пикселя области открытия/закрытия, $I[\cdot]$ — индикатор, σ — сигмоида, P — количество уровней в пирамиде, α_p — коэффициент для заданного уровня пирамиды. В предложенном нейросетевом методе были использованы те же параметры, что и в [27], увеличенные в 10 раз: $P = 6$, $\alpha_1 = 3.2$, $\alpha_2 = 0.8$, $\alpha_3 = 0.2$, $\alpha_4 = 0.1$, $\alpha_5 = 0.05$, $\alpha_6 = 0.025$. Для оптимизации был выбран метод Adam [34] со стандартными параметрами ($\beta_1 = 0.9$, $\beta_2 = 0.99$) и константным темпом обучения $lr = 10^{-4}$.

V. Результаты

A. Сравнение с базовым методом

Было проведено сравнение предложенного метода с базовым алгоритмом на наборе данных MPI-Sintel (Final). Precision-Recall кривая (Рис. 3) для всего набора данных показывает, что разработанный метод работает лучше базового. Значения F-меры представлены в Таблице III.

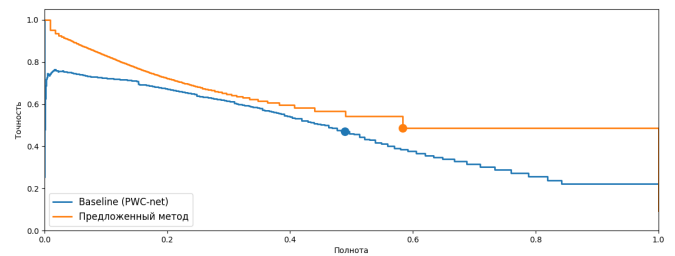


Рис. 3. Precision-Recall кривая, построенная на наборе данных MPI-Sintel (Final). Изменение порога для определения принадлежности пикселей области открытия/закрытия дает разные значения F-меры. Максимальное значение отмечено точкой.

B. Сравнение с другими методами

Было проведено сравнение разработанного алгоритма с оптимизационными методами на тестовом наборе данных MPI-Sintel (Final). Для оценки использовался тот же подход, что и в статье [35], где тестирование

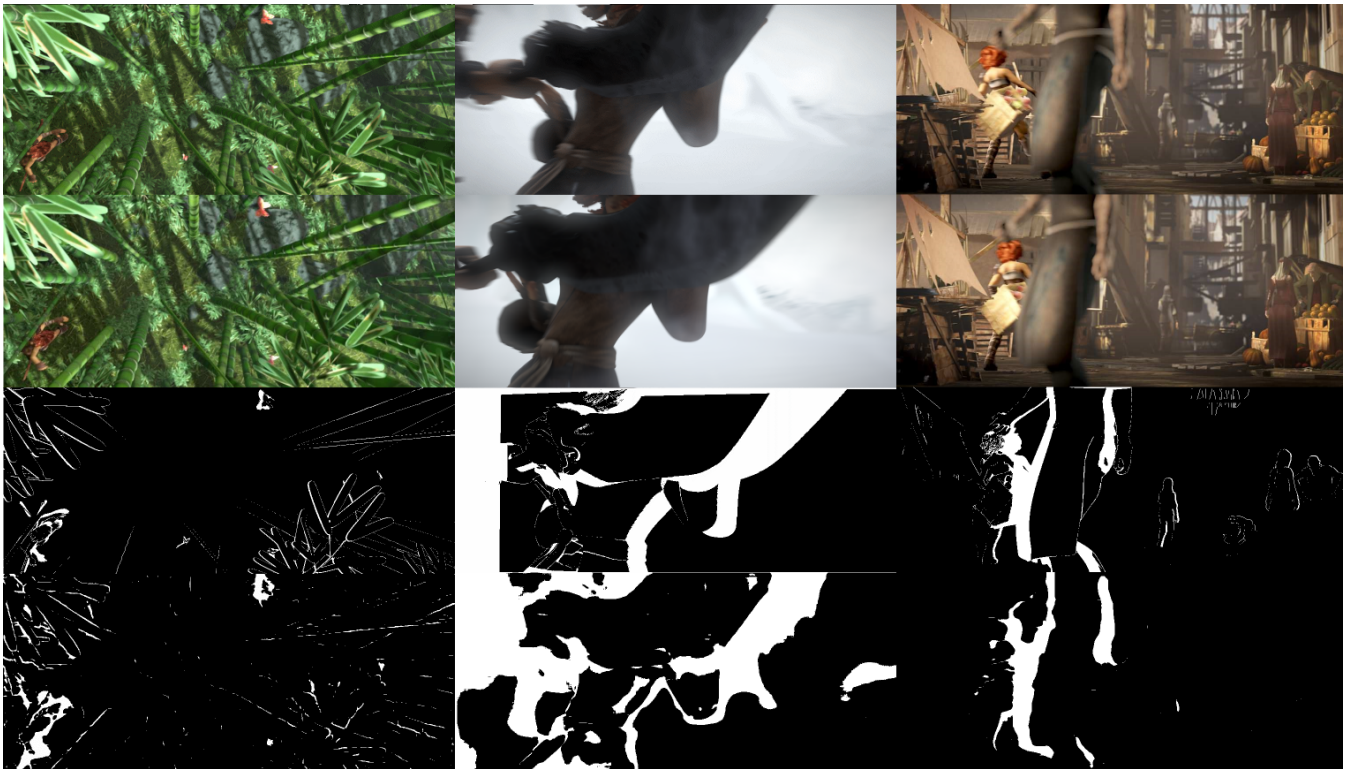


Рис. 4. Пример работы алгоритма на Sintel Final. Первая колонка — сцена *bamboo1*, вторая — *ambush2*, третья — *market2*. Первый ряд — первый кадр, второй ряд — второй кадр, третий ряд — истинные значения областей открытия/закрытия, четвертый ряд — результат предложенного метода.

Таблица III. Сравнение с базовым методом. F-мера посчитана по всем кадрам сцен.

Метод	Сцена					Весь набор данных
	alley1	bamboo1	bandage1	bandage2	market2	
Предложенный метод	0.320	0.530	0.470	0.462	0.566	0.531
Baseline(PWC-net)	0.281	0.547	0.450	0.439	0.557	0.479

Таблица IV. Сравнение с оптимизационными методами на пяти сценах из MPI-Sintel (Final). F-мера посчитана аналогично [35] по пяти кадрам из каждой сцены.

Метод	Сцена					Средняя F-мера
	alley1	bamboo1	bandage1	bandage2	market2	
Предложенный метод	0.11	0.32	0.50	0.35	0.71	0.40
SVU [36]	0.19	0.33	0.31	0.26	0.50	0.32
Estellers [35]	0.29	0.34	0.34	0.19	0.32	0.30
Ayvaci [15]	0.03	0.04	0.01	0.02	0.04	0.03
Alvarez [7]	0.26	0.36	0.22	0.15	0.38	0.27

проводилось на 5 сценах: *alley1*, *bamboo1*, *bandage1*, *bandage2* и *market2*. Итоговые результаты приведены в Таблице IV. Разработанный нейросетевой метод имеет в среднем лучшие результаты, но на сцене *alley1* сильно проигрывает остальным методам, так как в этой сцене очень узкие области открытия/закрытия шириной 1–2 пикселя, а предсказание карт областей открытия/закрытия осуществляется для разрешения в 4 раза меньше исходного.

Также было проведено сравнение с оптимизационными методами способом, предложенным в статье [11] (Таблица V). Нейросетевой метод SymmNet обучался и тестировался на MPI-Sintel 69 с использованием кросс-валидации с разбиением данных на 10 частей. Предложенный метод был дообучен на наборе данных MPI-Sintel 69. Результаты представлены в Таблице VI. На данный момент на наборе данных Sintel лучшим по F-мере нейросетевым методом для решения исследуемой задачи является метод на основе архитектуры FlowNet 2.0, описанный в статье [19]. Предложенный метод проигрывает данному методу

(Таблица VII), но имеет в 5 раз меньше параметров, что существенно уменьшает время обучения и значительно повышает скорость работы.

Таблица V. Сравнение с оптимизационными методами способом, предложенным в [11]. Oracle 69 и Global 69 включают все сцены из MPI-Sintel (Albedo, Clean и Final), но в Global подбирался оптимальный порог по всем сценам, а в Oracle — для каждой сцены отдельно. Oracle Final — аналогично, но только для MPI-Sintel (Final).

Метод	Oracle 69	Global 69	Oracle Final
Learning [14]	0.535	0.448	-
Depth order [16]	0.465	0.449	0.398
Local layers [17]	0.474	0.376	-
Sparse method [15]	0.310	0.259	0.258
Pérez-Rúa [11]	0.550	0.540	0.491
Предложенный метод	0.567	0.559	0.539

Таблица VI. Сравнение с нейросетевым методом SymmNet на наборе данных MPI Sintel 69 [11].

Метод	Oracle 69	Global 69
SymmNet [18]	0.665	0.642
Предложенный метод	0.702	0.693

VI. Заключение и дальнейшие исследования

В работе был предложен и реализован нейросетевой метод поиска областей открытия/закрытия

Таблица VII. Сравнение с нейросетевым методом FlowNet-CSSR-ft-sd.

Метод	Количество параметров	Время обучения	Global Clean	Global Final
FlowNet-CSSR-ft-sd [19]	117 млн.	12 д.	0.703	0.654
Предложенный метод	23 млн.	3 д.	0.562	0.531

в видеопоследовательностях, который при экспериментальном сравнении с оптимизационными и нейросетевыми аналогами на наборе данных MPI-Sintel показал значительное улучшение качества вычисляемых областей. За счет используемой в предложенном методе иерархической архитектуры PWC-Net нейросеть имеет небольшое количество параметров и высокую скорость работы и обучения, хотя по качеству он уступает методу, основанному на FlowNet 2.0.

Так как нейросетевые алгоритмы поиска оптического потока активно развиваются, возможно заимствование идей для совершенствования методов поиска областей открытия/закрытия. Возможными улучшениями являются использование более двух последовательных кадров и измененных расписаний обучения. Планируется применить данные улучшения в дальнейшей работе.

Список литературы

- Moving gradients: a path-based method for plausible image interpolation / Dhruv Mahajan, Fu-Chung Huang, Wojciech Matusik et al. // ACM Transactions on Graphics (TOG) / ACM. — Vol. 28. — 2009. — P. 42.
- Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation / Byeong-Doo Choi, Jong-Woo Han, Chang-Su Kim, Sung-Jea Ko // IEEE Transactions on Circuits and Systems for Video Technology. — 2007. — Vol. 17, no. 4. — P. 407–416.
- Cremers Daniel, Soatto Stefano. Motion competition: A variational approach to piecewise parametric motion segmentation // International Journal of Computer Vision. — 2005. — Vol. 62, no. 3. — P. 249–265.
- Chang Michael M, Tekalp A Murat, Sezan M Ibrahim. Simultaneous motion estimation and segmentation // IEEE transactions on image processing. — 1997. — Vol. 6, no. 9. — P. 1326–1333.
- Ascenso Joao, Brites Catarina, Pereira Fernando. Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding // 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services / Citeseer. — 2005. — P. 1–6.
- Puri A, Hang H-M, Schilling D. An efficient block-matching algorithm for motion-compensated coding // ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing / IEEE. — Vol. 12. — 1987. — P. 1063–1066.
- Symmetrical dense optical flow estimation with occlusions detection / Luis Alvarez, Rachid Deriche, Théo Papadopoulos, Javier Sánchez // International Journal of Computer Vision. — 2007. — Vol. 75, no. 3. — P. 371–385.
- Computing visual correspondence with occlusions via graph cuts : Rep. / Cornell University ; Executor: Vladimir Kolmogorov, Ramin Zabih : 2001.
- Strecha Christoph, Fransens Rik, Van Gool Luc. A probabilistic approach to large displacement optical flow and occlusion detection // International Workshop on Statistical Methods in Video Processing / Springer. — 2004. — P. 71–82.
- Hur Junhwa, Roth Stefan. Mirrorflow: Exploiting symmetries in joint optical flow and occlusion estimation // Proceedings of the IEEE International Conference on Computer Vision. — 2017. — P. 312–321.
- Determining occlusions from space and time image reconstructions / Juan-Manuel Pérez-Rúa, Tomas Crivelli, Patrick Boutheymy, Patrick Pérez // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — P. 1382–1391.
- Deepmatching: Hierarchical deformable dense matching / Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid // International Journal of Computer Vision. — 2016. — Vol. 120, no. 3. — P. 300–323.
- Lee Kyong Joon, Yun Il Dong. Occlusion detecting window matching scheme for optical flow estimation with discrete optimization // Pattern Recognition Letters. — 2017. — Vol. 89. — P. 73–80.
- Humayun Ahmad, Mac Aodha Oisín, Brostow Gabriel J. Learning to find occlusion regions // CVPR 2011 / IEEE. — 2011. — P. 2161–2168.
- Ayvaci Alper, Raptis Michalis, Soatto Stefano. Sparse occlusion detection with optical flow // International Journal of Computer Vision. — 2012. — Vol. 97, no. 3. — P. 322–338.
- Sun Deqing, Sudderth Erik B, Black Michael J. Layered image motion with explicit occlusions, temporal consistency, and depth ordering // Advances in Neural Information Processing Systems. — 2010. — P. 2226–2234.
- Sun Deqing, Liu Ce, Pfister Hanspeter. Local layering for joint motion estimation and occlusion detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2014. — P. 1098–1105.
- Li Ang, Yuan Zejian. Symmnet: A symmetric convolutional neural network for occlusion detection // arXiv preprint arXiv:1807.00959. — 2018.
- Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation / Eddy Ilg, Tonmoy Saikia, Margret Keuper, Thomas Brox // Proceedings of the European Conference on Computer Vision. — 2018. — P. 614–630.
- Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention / Springer. — 2015. — P. 234–241.
- Chaurasia Abhishek, Culurciello Eugenio. Linknet: Exploiting encoder representations for efficient semantic segmentation // 2017 IEEE Visual Communications and Image Processing (VCIP) / IEEE. — 2017. — P. 1–4.
- Bisenet: Bilateral segmentation network for real-time semantic segmentation / Changqian Yu, Jingbo Wang, Chao Peng et al. // Proceedings of the European Conference on Computer Vision. — 2018. — P. 325–341.
- FlowNet 2.0: Evolution of optical flow estimation with deep networks / Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 2462–2470.
- FlowNet: Learning optical flow with convolutional networks / Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg et al. // Proceedings of the IEEE International Conference on Computer Vision. — 2015. — P. 2758–2766.
- Xu Jia, Ranfil René, Koltun Vladlen. Accurate optical flow via direct cost volume processing // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 1289–1297.
- Epicflow: Edge-preserving interpolation of correspondences for optical flow / Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2015. — P. 1164–1172.
- Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume / Deqing Sun, Xiaodong Yang, Ming-Yu Liu, Jan Kautz // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — P. 8934–8943.
- Imagenet: A large-scale hierarchical image database / Jia Deng, Wei Dong, Richard Socher et al. — 2009.
- Deep residual learning for image recognition / Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — P. 770–778.
- A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation / Nikolaus Mayer, Eddy Ilg, Philip Hausser et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — P. 4040–4048.
- Egnal Geoffrey, Wildes Richard P. Detecting binocular half-occlusions: empirical comparisons of four approaches // Proceedings IEEE Conference on Computer Vision and Pattern Recognition / IEEE. — Vol. 2. — 2000. — P. 466–473.
- Empirical evaluation of rectified activations in convolutional network / Bing Xu, Naiyan Wang, Tianqi Chen, Mu Li // arXiv preprint arXiv:1505.00853. — 2015.
- Densely connected convolutional networks / Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q Weinberger // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 4700–4708.
- Kingma Diederik P, Ba Jimmy. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. — 2014.
- Estellers Virginia, Soatto Stefano. Detecting occlusions as an inverse problem // Journal of Mathematical Imaging and Vision. — 2016. — Vol. 54, no. 2. — P. 181–198.
- Ha Synh Viet-Uyen, Vu Tuan-Anh, Tran Ha Manh. An extended occlusion detection approach for video processing // REV Journal on Electronics and Communications. — 2019. — Vol. 8, no. 3–4.

A neural network approach for occlusion detection in video

M. Velikanov, A. Anzina, S. Lavrushkin, D. Vatolin

Abstract—Occlusions are a set of pixels, which are visible in a single frame of two sequential frames in a video. Finding occlusions is of great importance in the field of computer vision. Precise detection of occlusions will improve the accuracy of many video processing methods, such as: frame interpolation, optical flow calculation, color propagation etc. The majority of existing methods are based on optimization of an energy function, which is computationally expensive. It is also worth noting that accurate estimation of occlusions is hard with no information about movement between frames, and knowledge of occlusions during optical flow estimation allows the algorithm to avoid wrong correspondences between pixels of frames. Taking this into consideration we present a novel method of occlusion detection based on PWC-net, an optical flow calculation algorithm. The key idea is to construct a pyramid of features with different resolutions for frame processing. This way of processing originates from a common computer graphics technique and is widely adopted. We also performed a comparison of our method with 15 similar methods on the MPI-Sintel dataset.

Keywords—Deep learning, occlusions, neural networks, video processing.

References

- [1] Moving gradients: a path-based method for plausible image interpolation / Dhruv Mahajan, Fu-Chung Huang, Wojciech Matusik et al. // ACM Transactions on Graphics (TOG) / ACM. — Vol. 28. — 2009. — P. 42.
- [2] Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation / Byeong-Doo Choi, Jong-Woo Han, Chang-Su Kim, Sung-Jea Ko // IEEE Transactions on Circuits and Systems for Video Technology. — 2007. — Vol. 17, no. 4. — P. 407–416.
- [3] Cremers Daniel, Soatto Stefano. Motion competition: A variational approach to piecewise parametric motion segmentation // International Journal of Computer Vision. — 2005. — Vol. 62, no. 3. — P. 249–265.
- [4] Chang Michael M, Tekalp A Murat, Sezan M Ibrahim. Simultaneous motion estimation and segmentation // IEEE transactions on image processing. — 1997. — Vol. 6, no. 9. — P. 1326–1333.
- [5] Ascenso Joao, Brites Catarina, Pereira Fernando. Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding // 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services / Citeseer. — 2005. — P. 1–6.
- [6] Puri A, Hang H-M, Schilling D. An efficient block-matching algorithm for motion-compensated coding // ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing / IEEE. — Vol. 12. — 1987. — P. 1063–1066.
- [7] Symmetrical dense optical flow estimation with occlusions detection / Luis Alvarez, Rachid Deriche, Théo Papadopoulou, Javier Sánchez // International Journal of Computer Vision. — 2007. — Vol. 75, no. 3. — P. 371–385.
- [8] Computing visual correspondence with occlusions via graph cuts : Rep. / Cornell University ; Executor: Vladimir Kolmogorov, Ramin Zabih : 2001.
- [9] Strecha Christoph, Fransens Rik, Van Gool Luc. A probabilistic approach to large displacement optical flow and occlusion detection // International Workshop on Statistical Methods in Video Processing / Springer. — 2004. — P. 71–82.
- [10] Hur Junhwa, Roth Stefan. Mirrorflow: Exploiting symmetries in joint optical flow and occlusion estimation // Proceedings of the IEEE International Conference on Computer Vision. — 2017. — P. 312–321.
- [11] Determining occlusions from space and time image reconstructions / Juan-Manuel Pérez-Rúa, Tomas Crivelli, Patrick Bouthemy, Patrick Pérez // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — P. 1382–1391.
- [12] Deepmatching: Hierarchical deformable dense matching / Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid // International Journal of Computer Vision. — 2016. — Vol. 120, no. 3. — P. 300–323.
- [13] Lee Kyong Joon, Yun Il Dong. Occlusion detecting window matching scheme for optical flow estimation with discrete optimization // Pattern Recognition Letters. — 2017. — Vol. 89. — P. 73–80.
- [14] Humayun Ahmad, Mac Aodha Oisín, Brostow Gabriel J. Learning to find occlusion regions // CVPR 2011 / IEEE. — 2011. — P. 2161–2168.
- [15] Ayvaci Alper, Raptis Michalis, Soatto Stefano. Sparse occlusion detection with optical flow // International Journal of Computer Vision. — 2012. — Vol. 97, no. 3. — P. 322–338.
- [16] Sun Deqing, Sudderth Erik B, Black Michael J. Layered image motion with explicit occlusions, temporal consistency, and depth ordering // Advances in Neural Information Processing Systems. — 2010. — P. 2226–2234.
- [17] Sun Deqing, Liu Ce, Pfister Hanspeter. Local layering for joint motion estimation and occlusion detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2014. — P. 1098–1105.
- [18] Li Ang, Yuan Zejian. Symmnet: A symmetric convolutional neural network for occlusion detection // arXiv preprint arXiv:1807.00959. — 2018.
- [19] Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation / Eddy Ilg, Tomoy Saikia, Margret Keuper, Thomas Brox // Proceedings of the European Conference on Computer Vision. — 2018. — P. 614–630.
- [20] Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention / Springer. — 2015. — P. 234–241.
- [21] Chaurasia Abhishek, Culurciello Eugenio. Linknet: Exploiting encoder representations for efficient semantic segmentation // 2017 IEEE Visual Communications and Image Processing (VCIP) / IEEE. — 2017. — P. 1–4.
- [22] Bisenet: Bilateral segmentation network for real-time semantic segmentation / Changqian Yu, Jingbo Wang, Chao Peng et al. // Proceedings of the European Conference on Computer Vision. — 2018. — P. 325–341.
- [23] FlowNet 2.0: Evolution of optical flow estimation with deep networks / Eddy Ilg, Nikolaus Mayer, Tomoy Saikia et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 2462–2470.
- [24] FlowNet: Learning optical flow with convolutional networks / Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg et al. // Proceedings of the IEEE International Conference on Computer Vision. — 2015. — P. 2758–2766.
- [25] Xu Jia, Ranftl René, Koltun Vladlen. Accurate optical flow via direct cost volume processing // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 1289–1297.
- [26] Epicflow: Edge-preserving interpolation of correspondences for optical flow / Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2015. — P. 1164–1172.
- [27] Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume / Deqing Sun, Xiaodong Yang, Ming-Yu Liu, Jan Kautz // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — P. 8934–8943.
- [28] Imagenet: A large-scale hierarchical image database / Jia Deng, Wei Dong, Richard Socher et al. — 2009.
- [29] Deep residual learning for image recognition / Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun // Proceedings of the IEEE

- Conference on Computer Vision and Pattern Recognition. — 2016. — P. 770–778.
- [30] A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation / Nikolaus Mayer, Eddy Ilg, Philip Hausser et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — P. 4040–4048.
- [31] Egnal Geoffrey, Wildes Richard P. Detecting binocular half-occlusions: empirical comparisons of four approaches // Proceedings IEEE Conference on Computer Vision and Pattern Recognition / IEEE. — Vol. 2. — 2000. — P. 466–473.
- [32] Empirical evaluation of rectified activations in convolutional network / Bing Xu, Naiyan Wang, Tianqi Chen, Mu Li // arXiv preprint arXiv:1505.00853. — 2015.
- [33] Densely connected convolutional networks / Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q Weinberger // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 4700–4708.
- [34] Kingma Diederik P, Ba Jimmy. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. — 2014.
- [35] Estellers Virginia, Soatto Stefano. Detecting occlusions as an inverse problem // Journal of Mathematical Imaging and Vision. — 2016. — Vol. 54, no. 2. — P. 181–198.
- [36] Ha Synh Viet-Uyen, Vu Tuan-Anh, Tran Ha Manh. An extended occlusion detection approach for video processing // REV Journal on Electronics and Communications. — 2019. — Vol. 8, no. 3-4.