УДК 519.67, 519.688

С. Е. Ваганов

ПОЛИНОМИАЛЬНАЯ МОДЕЛЬ МЕЖКАДРОВОГО ДВИЖЕНИЯ В РЕШЕНИИ ЗАДАЧИ УТОЧНЕНИЯ ОПТИЧЕСКОГО ПОТОКА

Предложен алгоритм построения и уточнения оптического потока по нескольким кадрам видео посредством использования полиномиальных моделей межкадровых движений точек. Проведен сравнительный анализ точности оптических потоков, построенных по двум и трем кадрам. В работе показано, что использование квадратичных моделей межкадровых движений точек позволяет найти более точную аппроксимацию оптического потока по сравнению с двухкадровыми методами.

Ключевые слова: обработка видео, градиентный спуск, оптический поток, полиномиальная модель межкадрового движения.

An algorithm for constructing and refining an optical flow over several video frames by using polynomial models of inter-frame points motion is proposed. A comparative analysis of the accuracy of optical flows constructed in two and three frames is carried out. The paper shows that the use of quadratic models of inter-frame movements of points allows to find a more accurate approximation of the optical flow compared with the two-frame methods.

Key words: video processing, gradient descent, optical flow, polynomial based motion model.

Введение

Оптический поток используется при решении многих задач компьютерного зрения. Среди них, в частности, слежение за объектами, реконструкция трехмерной сцены по набору кадров, предсказание и интерполяция кадров видео, сегментация и многие другие. Классическим подходом к решению задачи нахождения оптического потока является метод Лукаса — Канаде [1, 3, 6], вычисляющий аппроксимацию векторного поля скоростей по паре кадров видео. Существует множество модификаций и обобщений данного подхода [2, 4, 5, 10]. В настоящей работе для проведения сравнительного анализа будет рассматриваться алгоритм [5], позволяющий строить по паре кадров видео оптические потоки, обладающие более низкой вариацией и более высокой точностью по сравнению с методом [6]. Также существуют нейросетевые и многокадровые методы [7, 9, 11], рассмотрение которых выходит за рамки настоящей работы.

Одним из недостатков подходов, выполняющих построение оптического потока по паре кадров [4, 5, 6, 10], является низкая точность прогнозирования промежуточных кадров. Оптический поток сопоставляет точкам первого кадра вещественный вектор сдвига на второй кадр. Координаты точки на промежуточном кадре вычисляются посредством его линейной интерполяции.

В силу нелинейной изменчивости скорости межкадрового движения точек, линейная аппроксимация приводит к существенным пространственным искажениям интерполированного кадра. Решение данной проблемы является актуальной задачей.

Предложенный в настоящей работе метод выполняет построение полиномиальных моделей, описывающих межкадровые траектории движения точек, то есть сопоставление моменту времени t координаты точки на плоскости. Производная модели по параметру t определяет векторное поле скоростей. В качестве примера рассмотрена квадратичная модель. По мнению автора, полиномиальные модели межкадровых движений более естественны с точки зрения решения задач, связанных с использованием оптического потока. Полиномиальные модели позволяют выполнить переход от набора оптических потоков к полиному и обратно. Предложенный метод может использоваться как для нахождения оптического потока, так и для уточнения существующего. Приведенные свойства полиномиальных моделей позволяют более гибко подходить к решению задач, основанных на использовании оптического потока.

В работе показано, что применение квадратичных моделей межкадровых движений точек позволяет существенно повысить точность оптического потока, построенного методом [5].

В п. 1 фиксируются некоторые общие понятия и обозначения, используемые в дальнейшей работе. В п. 2 представлено описание алгоритма построения квадратичных моделей межкадрового движения. В п. 3 сформулированы результаты сравнительного анализа предложенного метода с алгоритмом [5].

1. Основные понятия

Пусть f_0, f_1 и f_2 — последовательные кадры видео. Тогда задачу поиска оптического потока по трем кадрам можно сформулировать в следующем виде.

Моделью движения точки будем называть векторную функцию

$$V(t) = (x(t), y(t))^T,$$

сопоставляющую моменту времени t координаты точки на плоскости. В случае видео $V(0) = (x_0, y_0)$ — целочисленная точка с кадра f_0.

В общем виде задача нахождения модели движения точки может быть сформулирована в виде минимизации квадратичного функционала:

$$S = \sum_{t=1}^{N-1} \Big(f_t\big(x(t), y(t)\big) - f_0\big(x_0, y_0\big)\Big)^2 \to min. \tag{1}$$

Зафиксируем в качестве $V(t)$ полиномиальную модель движений

$$V(t) = \big(x_0,\, y_0\big)^T + \sum_{k=1}^{p} \big(a_k t^k,\, b_k t^k\big)^T,$$

где p соответствует степени многочлена. Тогда решение можно искать в виде:

$$\begin{cases} \dfrac{\partial S}{\partial a_1} = \ldots = \dfrac{\partial S}{\partial a_p} = 0, \\ \dfrac{\partial S}{\partial b_1} = \ldots = \dfrac{\partial S}{\partial b_p} = 0. \end{cases}$$

Неизвестные коэффициенты полиномов $\overline{z} = (a_1, ..., a_n, b_1, ..., b_n)$ будем находить методом градиентного спуска:

$$z_{n+1} = z_n - LR \cdot \frac{\nabla S}{\|\nabla S\|},$$

где LR определяет шаг смещения весовых параметров в направлении антиградиента ($LearningRate$), а $\|\nabla S\|$ — длина вектора градиента.

При работе с данной моделью число последовательных кадров видео должно быть не менее степени полинома p.

2. Алгоритм построения квадратичных моделей межкадрового движения

Далее в качестве $V(t)$ будет рассматриваться полином второй степени (случай трех кадров):

$$V(t) = \left(a_2 t^2 + a_1 t + x_0,\, b_2 t^2 + b_1 t + y_0\right)^T.$$

Решение будем искать посредством минимизации (1).

Для борьбы с осцилляциями можно требовать минимизации второй производной. Для этого к S необходимо добавить член

$$\lambda \frac{d^2 V(t)}{dt^2},$$

где коэффициент λ — константа, определяющая вклад регуляризации.

Посредством оптимизации предложенной модели можно получить аппроксимацию полиномиальных моделей и при нулевом начальном приближении (нулевые коэффициенты многочленов). Однако, при этом сходимость градиентного спуска будет весьма медленной. Для её улучшения предлагается использовать (в качестве начального приближения) коэффициенты квадратичных полиномов, вычисленные по паре оптических потоков, а именно с первого кадра на второй и с первого кадра на третий. Для нахождения оптического потока использовался метод [5].

Предлагаемый подход позволяет выполнять переходы от набора оптических потоков к полиномиальным моделям и в обратную сторону (из полиномиальных моделей строить оптические потоки).

Пусть

$$F_{ij}(x, y) = \left(FX_{ij}(x, y),\, FY_{ij}(x, y)\right)$$

— вектор сдвига точки (x, y) с кадра f_i на кадр f_j. Тогда для точки с координатами $p = (x_0, y_0)$ веса квадратичных полиномов определяются следующим образом:

$$\begin{aligned} a_1 &= 2 \cdot FX_{01}(p) - FX_{02}(p)/2, \\ a_2 &= -FX_{01}(p) + FX_{02}(p)/2, \\ b_1 &= 2 \cdot FY_{01}(p) - FY_{02}(p)/2, \\ b_2 &= -FY_{01}(p) + FY_{02}(p)/2. \end{aligned} \tag{2}$$

Расчет вектора сдвига $F_{01}(x, y)$ целочисленной точки кадра f_0 с координатами (x, y) на кадр f_1 можно выполнить следующим образом:

$$F_{01}(x, y) = V(1) - V(0) = \sum_{i=1}^{2} (a_i, b_i)^T, \tag{3}$$

где V — модель движения точки (x, y). Векторы движений для других кадров вычисляются аналогичным образом.

Для вычисления градиента использовалось численное дифференцирование. Вычисление аппроксимации частной производной функции S по аргументу z_i определяется следующим образом:

$$\frac{\partial S}{\partial z_i} = \frac{S(z_1, ..., z_i + \varepsilon, ..., z_n) - S(z_1, ..., z_i - \varepsilon, ..., z_n)}{2\varepsilon},$$

где $\overline{z} = (a_1, ..., a_n, b_1, ..., b_n)$ — вектор неизвестных коэффициентов полиномиальной модели движения.

Уточнение оптического потока. Алгоритм уточнения оптического потока включает в себя последовательное выполнение следующих шагов.

1. Вычислить начальное приближение посредством (2). Установить $LR = 1$.

2. Выполнить шаг градиентного спуска с $\varepsilon = LR$.

3. Если качество потока ухудшилось, то выполнить замену текущих весов полиномов на коэффициенты с предыдущей итерации и положить $LR = LR/10$.

4. Если число прошедших итераций меньше N, то увеличить счетчик итераций и перейти к шагу 2.

5. Выполнить расчет оптического потока с первого кадра на второй посредством (3).

3. Результаты

Расчет коэффициентов полиномиальных моделей осуществлялся по набору первых троек видео-кадров [8]. Все кадры были уменьшены до размера 320×240 пикселей. Начальное приближение коэффициентов полиномов вычислялось по паре оптических потоков, найденных методом [5]. Для нахождения оптических потоков изображения были размыты фильтром Гаусса с размером ядра 3×3. Число итераций алгоритма $N = 32$.

Для проведения сравнительного анализа предложенного метода уточнения оптического потока с алгоритмом [5] (для первых пар кадров [8]) использовались следующие оценки.

Средняя длина вектора сдвига, найденная по оптическому потоку F для кадров размера $m \times n$:

$$MD(F) = \sqrt{\frac{1}{mn}\sum_{x=1}^{n}\sum_{y=1}^{m} \|F(x, y)\|^2}.$$

Корень из среднеквадратичного отклонения интенсивностей точек кадра f_0 и их образов на кадре f_1, найденных посредством оптического потока F:

$$RMSE(f_0, f_1, F) = \sqrt{\frac{1}{mn}\sum_{x=1}^{n}\sum_{y=1}^{m} \Big(f_1\big((x, y) + F(x, y)\big) - f_0\big(x, y\big)\Big)^2},$$

где значения f_1 в нецелых точках вычисляются посредством билинейной интерполяции.

Качество потока SPF (построенного по квадратичному полиному, Square Polynomial Flow), определяемое процентным соотношением величины $RMSE$ предложенного метода и алгоритма FF (Farnebäck Flow) [5]:

$$Q = 100 \cdot \frac{RMSE(f_0, f_1, SPF)}{RMSE(f_0, f_1, FF)}.$$

Оценки качества и средней длины оптических потоков на наборе кадров [8]

Видео	$MD(FF)$	R_{FF}	R_{SPF}	$Q, \%$
bear	0,861	3,131	2,179	69,587
bike-packing	0,654	2,905	2,060	70,911
blackswan	3,090	6,376	5,026	78,825
bmx-bumps	10,434	8,831	6,160	69,753
bmx-trees	9,104	13,081	10,541	80,585
boat	0,979	2,339	1,694	72,426
boxing-fisheye	3,208	7,984	5,910	74,025
breakdance	3,243	8,779	6,152	70,075
breakdance-flare	3,291	7,488	5,445	72,726
bus	2,411	3,354	2,507	74,733
camel	2,537	5,701	4,215	73,936
car-roundabout	7,612	12,196	7,637	62,624
car-shadow	2,991	5,193	4,135	79,633
car-turn	0,197	1,596	1,175	73,618
cat-girl	15,882	18,835	14,606	77,549
classic-car	10,233	11,429	8,449	73,926
color-run	4,400	6,102	3,999	65,530
cows	1,785	3,570	2,648	74,175
crossing	0,751	2,855	1,738	60,896
dance-jump	4,850	7,567	5,453	72,059
dance-twirl	2,116	4,708	3,405	72,310
dancing	3,461	14,316	9,881	69,018
disc-jockey	5,890	5,706	4,076	71,437
dog	5,485	5,374	3,360	62,522
dog-agility	6,513	18,303	14,361	78,466
Среднее	**4,479**	**7,509**	**5,473**	**72,054**

В приведенной таблице

$$R_{FF} = RMSE(f_0, f_1, FF), \quad R_{SPF} = RMSE(f_0, f_1, SPF),$$

где f_0 и f_1 — первая пара кадров видео. В последней строке представлены значения усредненных оценок.

Результаты численного эксперимента показывают, что минимизация (1) позволяет повысить точность исходных оптических потоков. При 32 итерациях предложенного алгоритма точность оптического потока с первого кадра на второй возрастает (в среднем) более чем на 25 %.

Предложенный в работе алгоритм может использоваться как для нахождения оптического потока по некоторому начальному приближению, так и для уточнения существующего. Использование пирамидального подхода позволит улучшить сходимость предложенного метода для случая нахождения коэффициентов с нулевым начальным приближением. Построение моделей по большему числу кадров позволит улучшить полученный в настоящей работе результат. Модификации предложенного алгоритма могут использоваться для решения задач интерполяции и сегментации кадров видео.

Библиографический список

1. *Ваганов С. Е., Хашин С. И.* Сравнение эффективности различных версий метода Лукаса — Канаде // Вестник Ивановского государственного университета. Сер.: Естественные, общественные науки. 2017. Вып. 2. С. 63—70.
2. *Хашин С. И.* Аффинная версия алгоритма Лукаса — Канады // Доклады Всероссийской конференции ММРО-13. М.: МАКС Пресс, 2011. С. 459—462.
3. *Baker S., Gross R., Matthews I.* Lucas — Kanade 20 years on: a unifying framework // Int. J. Computer Vision. 2002. Vol. 56. P. 111—122.
4. *Brox T., Bruhn T., Papenberg N., Weickert J.* High accuracy optical flow estimation based on a theory for warping // *Pajdla T., Matas J.* (eds) Computer Vision — ECCV 2004. Berlin, Heidelberg: Springer, 2004. P. 25—36. (Lecture Notes in Computer Science. Vol. 3024.)
5. *Farnebäck G.* Two-frame motion estimation based on polynomial expansion // *Bigun J., Gustavsson T.* (eds) Image Analysis. SCIA 2003. Berlin, Heidelberg: Springer, 2003. P. 363–370. (Lecture Notes in Computer Science. Vol. 2749.)
6. *Lucas B. D., Kanade T.* An iterative image registration technique with an application to stereo vision // Proc. of Imaging Understanding Workshop. 1981. P. 121—130.
7. Optical flow evaluation results : сайт. URL: http://vision.middlebury.edu/flow/eval/results/results-i1.php (дата обращения: 02.04.2019).
8. *Perazzi F., Pont-Tuset J., McWilliams B., Van Gool L., Gross M., Sorkine-Hornung A.* A benchmark dataset and evaluation methodology for video object segmentation // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. URL: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Perazzi_A_Benchmark_Dataset_CVPR_2016_paper.pdf (дата обращения: 02.04.2019).
9. *Teney D. Hebert M.* Learning to extract motion from videos in convolutional neural networks // arXiv:1601.07532. URL: https://arxiv.org/abs/1601.07532 (дата обращения: 02.04.2019).
10. *Wedel A., Pock T., Braun J., Franke J., Cremers D.* Duality TV-L1 flow with fundamental matrix prior // 23rd International Conference Image and Vision Computing. New Zealand, 2008. URL: https://ieeexplore.ieee.org/abstract/document/4762119 (дата обращения: 02.04.2019).
11. *Weinzaepfel P., Revaud J., Harchaoui Z., Schmid C.* DeepFlow: large displacement optical flow with deep matching // IEEE International Conference on Computer Vision (ICCV). 2013. URL: https://www.cv-foundation.org/openaccess/content_iccv_2013/papers/Weinzaepfel_DeepFlow_Large_Displacement_2013_ICCV_paper.pdf (дата обращения: 02.04.2019).