

ПОМЕХОУСТОЙЧИВЫЙ МЕТОД ПОСТРОЕНИЯ ОПТИЧЕСКОГО ПОТОКА ПО НЕСКОЛЬКИМ КАДРАМ

© 2010 г. В. А. Козлов

Санкт-Петербургский государственный университет информационных технологий,
механики и оптики, Санкт-Петербург

E-mail: vadikus@mail.ru

Предложен метод построения оптического потока с использованием более двух последовательных кадров. За счет анализа более двух кадров метод высоко устойчив к помехам и отслеживает даже медленные движения, а использование каскадов делает его быстрым. В итоге метод способен строить качественный оптический поток в реальном времени даже по запущенным видеозаписям с бытовых камер независимо от снимаемой сцены.

Ключевые слова: оптический поток, анализ движения, временное сглаживание, анализ во времени, несколько кадров, каскады.

Коды OCIS: 110.4850, 110.4153, 110.4280

Поступила в редакцию 30.04.2010

Введение

Соседние кадры видеопоследовательности, как правило, содержат изображения одних и тех же движущихся объектов в разное время. Оптический поток – это информация о том, куда смещается каждый пиксел на следующем кадре. Другими словами, оптический поток – это поле векторов, с помощью которого можно получить следующий кадр из предыдущего.

Построение оптического потока является одним из базовых элементов большинства алгоритмов анализа движения. Соответственно, от точности и адекватности построенного оптического потока зависит качество результата всего алгоритма. Низкое быстродействие построения оптического потока может значительно сказаться на общей скорости работы всего алгоритма и даже сделать невозможным применение его в реальном времени. Многие алгоритмы построения оптического потока качественно работают только на определенных типах сцен или требуют очень тонкой настройки параметров для каждой определенной сцены [1–6]. А при появлении естественного шума камеры, например при съемках в затемненных условиях, некоторые алгоритмы полностью перестают работать или дают плохие результаты, указывая перемещения случайного шума, при этом теряя перемещения слабоконтрастных объектов.

Одной из причин описанных выше проблем является построение этими алгоритмами оптического потока только по двум соседним кадрам видеопоследовательности, что уже само по себе не соответствует физиологии человеческого восприятия движения [7–12]. Как показали исследования в смежных областях распознавания изображений и анализа движения, достаточно часто более адекватные результаты дают алгоритмы, основанные на тех же принципах, на которых человеческий глаз совместно с мозгом воспринимает видимый мир.

Как известно, человек воспринимает движение только по большому числу изменений положения объекта в одном направлении [9–11]. В человеческом мозгу представление о движении объекта не формируется только лишь по двум изображениям или по случайным скачкам объекта. Кроме того, скорость движения должна лежать в определенном диапазоне – очень медленные или очень быстрые движения тоже не фиксируются [12]. Таким образом, в человеке каким-то образом изначально заложено распознавание движения по ряду кадров. Тем не менее, применяемые сейчас алгоритмы используют для анализа только два последовательных кадра при построении оптического потока. Вероятно, это было связано с ограничениями вычислительных модулей прошлых лет, но современный уровень технического прогресса позволяет ис-

пользовать требовательные к ресурсам алгоритмы для достижения более качественного построения оптического потока.

Обычно человеческий глаз видит не менее 60 кадров в секунду, но воспринимается меньшее число кадров. Например, расслабленный человек, идущий по улице, видит как бы слизанные изображения по 10 кадров в секунду [7, 8, 12], и только в случае экстремальных ситуаций мозг пытается воспринять и проанализировать движение максимально детально. Но даже в процессе такого облегченного восприятия реальности человек достаточно четко фиксирует движения всех объектов.

При изучении новой сцены человек сначала выделяет и распознает составляющие ее объекты, часто основываясь на их перемещениях, а только потом рассматривает детали объектов в течение более длительного времени. То есть распознавание движения происходит раньше, чем детализация предметов [9–11]. Как известно, если объект наблюдения двигался очень быстро и исчез из поля зрения за короткое время, то человек способен описать траекторию движения и, возможно, форму объекта, но не способен заметить никаких деталей [12].

Принимая во внимание все отмеченные здесь аспекты человеческого зрения, был разработан метод построения оптического потока по нескольким кадрам, который до определенной степени лояльно относится к шумам и использует историю кадров для достижения качественного результата, не зависящего от сцены.

Сглаживание шумов

Для сглаживания шумов будем разрабатывать алгоритм анализа движения по аналогии с человеческим зрением, а именно будем рассматривать не каждый кадр, а сглаженные кадры, которые получены усреднением нескольких соседних кадров.

Яркость каждого пикселя в цифровой фотографии содержит шумовую составляющую. Уровень шумов, естественно, зависит от качества используемых фотоприемников, однако для многих практических задач существенны ограничения на стоимость используемого оборудования, а также и на условия съемки. При этом в ряде случаев отношение сигнал/шум может достигать весьма малых значений.

При усреднении пикселов одной и той же точки на кадрах в разные моменты времени в подавляющем большинстве случаев рассматри-

вается оцифрованная точка приблизительно одной яркости, т. е. усредняется истинное значение яркости, а шум уменьшается, в отличие от усреднения соседних пикселов в одном кадре, у которых разница яркостей больше. Таким образом, временное сглаживание оцифрованных кадров более эффективно, чем пространственное сглаживание. К тому же, можно отметить, что небольшая слизанность кадров от временного сглаживания хорошо устраняет шумы и не влияет на построение оптического потока, потому что движение должно продолжаться несколько кадров подряд в примерно одном направлении, иначе для алгоритма анализа движения, как и для человека, это не будет являться движением.

А ведь это очень важная деталь: человек сначала видит, что объект изменяет свое положение несколько раз, и только после этого формирует представление о движении. Будем поступать аналогичным образом.

Очередь кадров

Как отмечено выше, первым шагом разрабатываемого метода является временное сглаживание изображений для уменьшения шумов. Далее алгоритм анализирует полученные кадры в совокупности.

Поступающие на вход кадры буферизуются в очереди некоторой постоянной длины. Оптический поток рассчитывается для самого старого кадра в очереди. При вычислении, куда сдвигается каждый пиксель на следующем кадре, проводится поиск блоков, похожих на блок, окружающий пиксель. Причем поиск проходит не только в следующем кадре, но и во всех кадрах, помещенных в буфер. Если самый подходящий блок нашелся через несколько кадров вперед, то проставляется такое же смещение и для всех промежуточных кадров. Это позволяет выявить даже медленные движения. На рис. 1 приведен самый простой пример медленного движения. Вообще говоря, это позволяет хорошо выявлять движения любой скорости, а не только те, которые происходят со скоростью в целое число пикселов за время, равное расстоянию между кадрами, как это делают традиционные

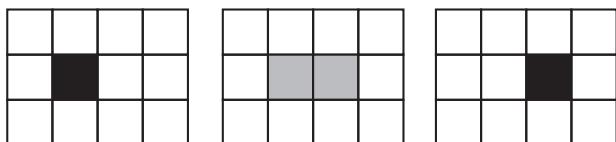


Рис. 1. Пример медленного движения.

методы. Таким образом, метод становится независимым от числа кадров в секунду, которое подается ему на вход, и от размеров кадров. Для понижения минимальной скорости выявляемого движения потребуется только увеличить размер очереди.

После вычисления оптического потока кадр выходит из очереди, а в очередь добавляется новый входящий кадр. Таким образом, вычисление оптического потока осуществляется с небольшой задержкой, равной размеру очереди, но без этого метод не будет обладать полезным свойством – определять медленные движения, а это потребуется на следующем этапе.

Каскады

Вспомним, что человек первоначально формирует представление о движении объекта по его образу (контуру и ключевым точкам), без рассмотрения деталей самого объекта. Метод будет имитировать аналогичные действия, используя каскады. А именно, оптический поток будет строиться для кадра в малом разрешении, а потом постепенно детализироваться.

Для каждого кадра в очереди будет строиться каскад, т. е. пирамида изображений, где каждое следующее изображение меньше предыдущего в два раза по ширине и высоте. Таким образом уменьшаем разрешение, пока ширина и высота делятся пополам. Например, для часто встречающегося в быту размера кадра 640×480 получим каскад из шести уровней, самое маленькое изображение – 20×15 . В подавляющем большинстве случаев каскад будет содержать достаточно изображений, потому что все стандартные разрешения видеокадров, получаемых после оцифровки, состоят из целого числа блоков 8×8 пикселов, т. е. каскад будет состоять минимум из четырех уровней. В случае использования нестандартных размеров кадра следует выполнить предварительную обрезку краев, чтобы в результате значения ширины и высоты кадра делились на восемь.

Сначала строится оптический поток для самого низкого разрешения, при этом, как и раньше, для поиска блоков проверяется не только соседний кадр, но и все кадры такого же разрешения в буфере. Далее оптический поток копируется на один уровень выше по каскаду, умножая смещения на два, ищется оптический поток на новом уровне. Похожие блоки ищутся вокруг текущей рассматриваемой точки и вокруг смещения, вычисленного на предыдущем уровне.

Более подходящий блок задаст новое смещение, которое будет передано на следующий уровень. Так происходит до тех пор, пока не будет достигнута вершина каскада, где и будет получен искомый оптический поток.

Отслеживание медленных скоростей за счет просмотра всего буфера дает возможность построения адекватного оптического потока на нижних уровнях каскада, на которых все движения медленные, потому что размер изображения мал. Большинство движений там будут похожи на пример, приведенный на рис. 1.

Итоговый метод

Подытоживая описание метода, приведенное выше, получаем следующий, кратко записанный, алгоритм.

Перед попаданием в основную очередь обработки приходящие кадры сглаживаются. Для каждого сглаженного кадра строится каскад и помещается в конец очереди каскадов.

Главный цикл обработки:

- Строится оптический поток для кадра из первого в очереди каскада.
- Процесс начинается с самого малого кадра в каскаде.
- Строится оптический поток для текущего уровня каскада.
- Переносится результат на уровень выше.
- Пока не будет достигнут исходный кадр, переход к пункту a2).
- На выход передается кадр и его оптический поток.
- Из очереди удаляется текущий анализируемый каскад.
- В конец очереди добавляется новый каскад.
- Переход к пункту a).

На рис. 2 представлена схема пунктов от a1 до a4. Первое – поиск в истории, второе – перенос на следующий уровень каскада.

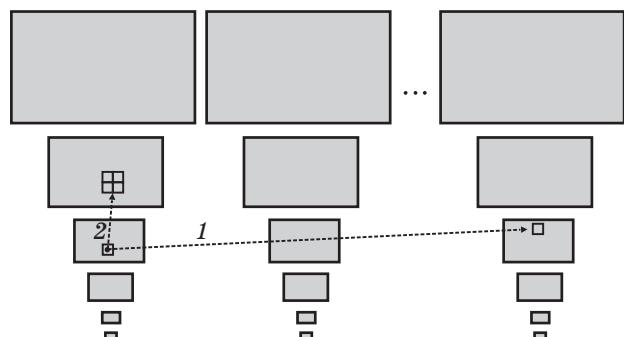


Рис. 2. Базовая схема метода.

Детали реализации

В подавляющем большинстве случаев для сглаживания достаточно вычислять полусумму двух кадров. Только если видеоряд снят с очень сильным шумом, следует усреднять три или четыре кадра. Большее число кадров для усреднения практически не дает эффекта.

Размер очереди обратно пропорционален преобладающим скоростям, которые присутствуют на видеопоследовательности. Обычно достаточно задать размер очереди в пять кадров. Очень большая очередь увеличивает нагрузку на процессор и требует больший объем оперативной памяти. Маленькая очередь не даст хорошего вычисления оптического потока для медленно двигающихся объектов.

Применение каскадов дает возможность использовать самые малые блоки поиска – 3×3 , что существенно уменьшает вычислительную сложность метода по сравнению со многими другими, которые, например, используют блоки 8×8 . Кстати, само применение каскадов увеличивает анализируемую площадь изображений всего в два раза.

В качестве метрики при сравнении блоков разумно использовать сумму модулей разности значений яркости или модулей разности производных значений яркости по пикселам внутри блока. Но можно использовать и любую другую подобную метрику.

У метода есть настроочные параметры, но в отличие от остальных методов, при любом значении параметров оптический поток будет вычислен адекватно в любой сцене. Настройка в данном методе менее критична, чем в других методах, и лишь немного улучшает качество результатов.

Экспериментальная проверка

Рассмотрим ряд известных классических алгоритмов построения оптического потока [1–3], которые реализованы в общедоступной библиотеке OpenCV. А именно сравним описываемый метод (будем его идентифицировать как KVA) с алгоритмами Хорна–Шанка (*Horn–Schunck – HS*), Лукаса–Канадэ (*Lucas–Kanade – LK*) и сопоставление блоков (*Block Matching – BM*).

Рассмотрим видеоряд, один кадр из которого представлен на рис. 3. Это запись с камеры на-



Рис. 3. Пример кадра видеоряда.

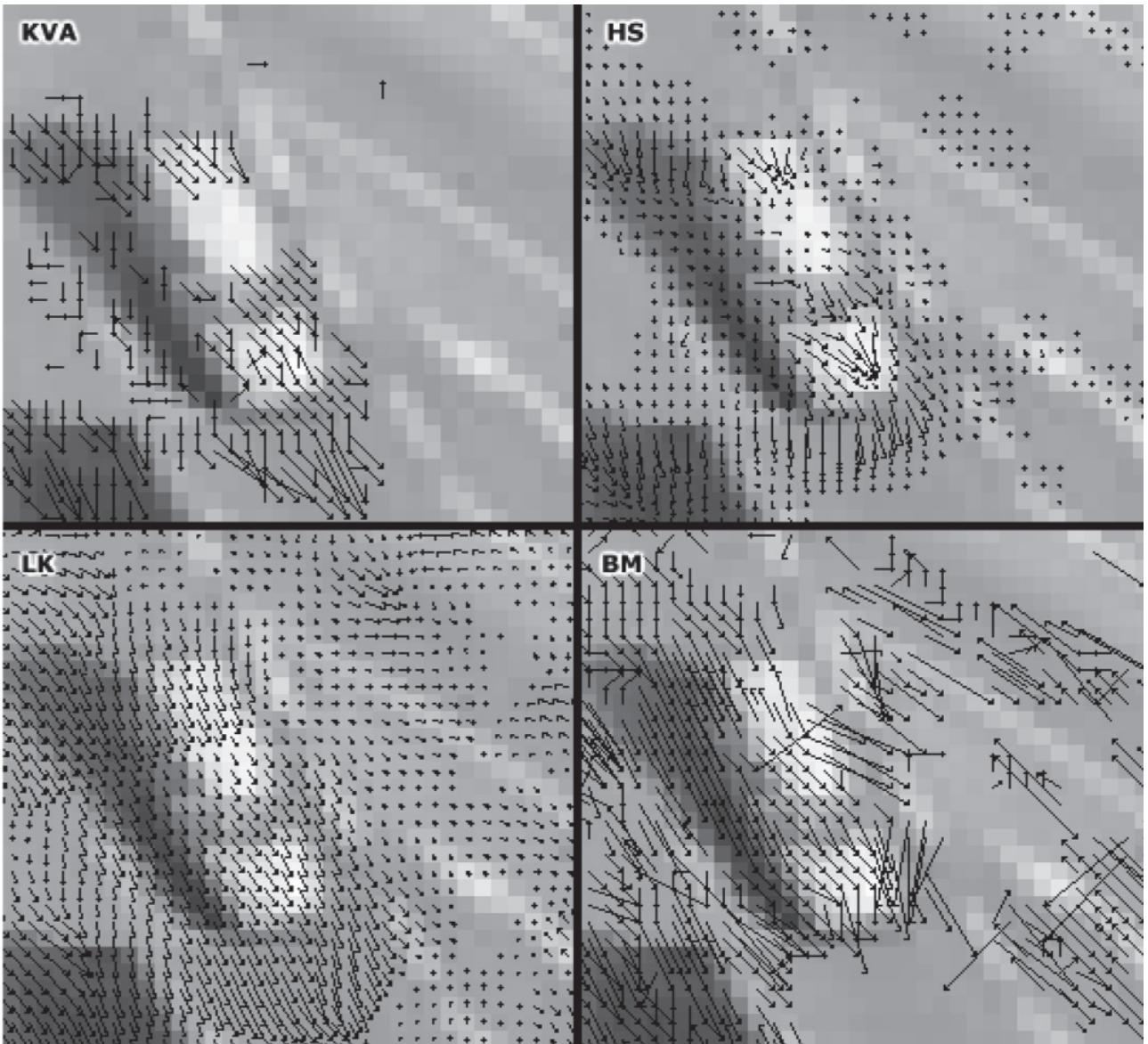


Рис. 4. Пример построения оптических потоков для сравнения.

блудения за перекрестком. Машины движутся вправо вниз. На записи присутствуют заметные естественные шумы.

На рис. 4 приведен результат построения оптического потока для одного и того же кадра всеми сравниваемыми алгоритмами. Для наглядности представлена только увеличенная часть кадра, которая обведена рамкой на рис. 3. Отсутствие стрелки оптического потока соответствует нулевому смещению пикселя. Алгоритмы HS и LK полагают, что оптический поток непрерывный, поэтому результат их работы выглядит гладко и красиво, но он не совсем соответствует действительности, практически никак не выделяется граница движущегося объекта. Алгоритм BM

просто ищет похожие блоки в соседнем кадре, поэтому его результат ближе к действительности, но наличие естественных шумов его сильно снижает. Все рассмотренные классические алгоритмы показали плохую устойчивость к шумам. С одной стороны, они выполнили задачу правильно – указали куда смещаются пиксели, но часто от алгоритма построения оптического потока требуется отслеживание движения именно объектов, а не шумов. С этой точки зрения, описанный в статье алгоритм показал хороший результат: он смог определить, что движения асфальта на самом деле нет, и проигнорировал шумы, поскольку он анализировал несколько кадров.

Далее сравним производительность алгоритмов. Самым быстрым оказался алгоритм LK. В пять раз дольше оптический поток вычисляет алгоритм HS. В шесть раз дольше работает описываемый в статье алгоритм, чем LK. А алгоритм BM работает в 12 раз дольше, чем LK.

Хотя алгоритм LK и самый быстрый, но его результат и самый некорректный, потому что он показывает, что движутся почти все пиксели в кадре. Описываемый в статье алгоритм оказался почти схожим по скорости с HS и в два раза более быстрым, чем “лобовой” алгоритм BM, при этом описываемый алгоритм строит самый качественный оптический поток.

Заключение

Аналогия с человеческим зрением помогла получить метод для построения оптического потока, качество результата которого в целом выше, чем у традиционных методов.

Помехоустойчивость и точность за счет анализа нескольких кадров позволяют применять метод для построения оптического потока даже по зашумленной видеопоследовательности с бытовых камер. При этом метод обладает хорошим быстродействием, что дает возможность применять его в системах, где требуется решать задачи анализа движения в реальном времени. Простота реализации позволяет даже заменять другие методы в рабочих системах с минимальными сторонними эффектами.

Основными недостатками предложенного метода являются требование большого объема памяти для хранения очереди каскадов и небольшая временная задержка при оценивании оптического потока. Тогда как в большинстве современных систем существенные ограничения на оперативную память отсутствуют.

ЛИТЕРАТУРА

1. Horn B.K.P., Schunck B.G. Determining optical flow // Artificial Intelligence. 1981. V. 17. P. 185–203.
2. Lucas B.D., Kanade T. An iterative image registration technique with an application to stereo vision // Proc. of Imaging understanding workshop. Washington, DC, 23 apr. 1981. P. 121–130.
3. Bouguet J.Y. Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm // OpenCV Documentation, Microprocessor Research Labs, Intel Corp, 2000.
4. Proesmans M., Van Gool L., Pauwels E., Oosterlinck A. Determination of optical flow and its Discontinuities using non-linear diffusion // In 3rd European Conf. on Computer Vision. Stockholm, Sweden, 2–6 May 1994. V. 2. P. 295–304.
5. Camus T. Real-Time Quantized Optical Flow // Journal of Real-Time Imaging. 1997. V. 3. P. 71–86.
6. Bruhn A., Weickert J., Schnorr C. Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic // International Journal of Computer Vision 61. 2005. V. 3. P. 211–231.
7. Хьюбел Д. Глаз, мозг, зрение. М.: Мир, 1990. 239 с.
8. Белопольский В.И. Взор человека: механизмы, модели, функции. М.: Институт психологии РАН, 2007. 415 с.
9. Логвиненко А.Д. Зрительное восприятие пространства. М.: МГУ, 1981. 223 с.
10. Грегори Р.Л. Глаз и мозг. Психология зрительного восприятия. М.: Прогресс, 1988. 272 с.
11. Креч Д., Крачфилд Р., Ливсон Н. Восприятие движения и времени // Хрестоматия по ощущению и восприятию / под ред. Гиппенрейтер Ю.Б., Михалевской М.Б. М.: Изд-во МГУ, 1975. С. 371–385.
12. Breitmeyer B.G., Ganz L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing // Psychol Rev. 1976. V. 83. P. 1–36.