

Использование карт глубины при восстановлении фона в видеопоследовательностях

Антон Зачесов,

Институт прикладной математики им. М.В. Келдыша РАН

azachesov@graphics.cs.msu.ru

Михаил Ерофеев, Дмитрий Ватолин

Московский государственный университет им. М. В. Ломоносова

merofeev@graphics.cs.msu.ru, dmitriy@graphics.cs.msu.ru

Аннотация. В данной статье предлагается метод автоматического восстановления неизвестного фона в видеопоследовательности для использования его при построении ракурсов стерео или многоракурсного видео. В качестве основы для определения областей неизвестного фона используется последовательность карт глубины для каждого кадра исходной видеопоследовательности. Предлагается алгоритм сопоставления фона из разных кадров видеопоследовательности, использующий карту глубины при анализе движения между кадрами. Предложенный метод кроме восстановленного фона для каждого кадра исходной видеопоследовательности также строит карту глубины этого восстановленного фона. Приведены результаты работы данного метода на видеопоследовательности со сложным движением фона и камеры.

Ключевые слова: восстановление фона, реконструкция видеопоследовательностей, оценка движения.

1 Введение

Рассмотрим задачу построения ракурсов стерео или многоракурсного видео из двумерного видео с использованием карты глубины для каждого кадра этого видео. Пусть для построения ракурсов используется сдвиг пикселей кадра вдоль горизонтальной оси вправо и влево относительно их первоначального положения, а величина сдвига пропорциональна значению интенсивности пикселя с такими же координатами на карте глубины. Одной из подзадач, возникающих при решении данной задачи, является восстановление фона, открывающегося при сдвиге пикселей за объектами переднего плана, расположенными ближе к наблюдателю (далее – задача восстановления фона или просто восстановление фона). Для естественного восприятия наблюдателем получаемой трехмерной сцены необходимо, чтобы эти области содержали информацию о «настоящем» фоне сцены, то есть фоне, который действительно присутствовал в сцене в момент ее съемки.

В ряде случаев, например, если сцена статична, информация о фоне может быть утеряна. Если же сцена не является статичной (объекты в

сцене движутся относительно друг друга или изменяется точка, с которой происходит захват кадров видео – далее движение камеры), фон, требующийся для построения ракурсов для текущего кадра, присутствует в других кадрах видео. В этом случае для восстановления фона необходимо перенести доступную информацию между кадрами видео. Для этого необходимо, во-первых, для каждого кадра определить области, нуждающиеся в восстановлении фона (далее эта область также называется маской восстановления), и, во-вторых, для каждого кадра видео найти в других кадрах и перенести соответствующие области фона. Этот частный случай задачи восстановления фона для построения ракурсов назовем восстановлением фона из движения в сцене. Из формулировки задачи следует, что от метода требуется восстановление небольших по площади областей (открывающиеся при сдвиге по карте глубины области имеют небольшую площадь), должны отсутствовать визуально заметные ошибки восстановления (в этом случае будет нарушена корректность отображения исходной трехмерной сцены на построенных ракурсах). Поскольку для построения ракурсов используется карта глубины, то она же может использоваться в процессе восстановления фона. Это налагает дополнительное требование на метод, используемый для восстановления фона – кроме непосредственно кадра с восстановленным фоном требуется построить карту глубины фона для этого кадра.

2 Обзор существующих методов

Известные методы реконструкции фона применяются для решения задачи удаления нежелательных объектов из видеопоследовательности (попавших в кадр во время съемки видеопоследовательности) или добавленных в видеопоследовательность в процессе ее постобработки, а также для восстановления поврежденных видеопоследовательностей, в которых утрачена информация о некоторых областях кадров. Первая задача часто решается для видеопоследовательностей, полученных с камер слежения. Для таких видеопоследовательностей характерны отсутствие движения камеры и неподвижный фон. Для решения могут применяться подходы на основе вычитания фона [Piccardi 2004; Crivelli et al., 2008]. В простейшем случае предполагается, что фон между кадрами видеопоследовательности либо не изменяется, либо не изменяется его структура (могут изменяться условия освещенности сцены). Для уточнения результата может применяться анализ движения в видео для отбрасывания из рассмотрения областей, содержащих движущиеся объекты.

Подобные методы позволяют построить точную аппроксимацию фона в случае неподвижной камеры, однако не применимы в случае заметных изменений структуры фона (вызванные движением камеры или самого фона). Это не позволяет использовать их для реконструкции фона в произвольной видеопоследовательности.

В случае произвольного движения для корректного восстановления фона требуется информация о том, какие области кадра принадлежат области фона, а какие принадлежат объектам переднего плана. Для решения задачи для видеопоследовательностей с произвольным движением в [Sheikh, 2009; Воронин и др., 2012] вводится модель движения фона, после чего применяется метод, аналогичный вычитанию фона, учитывающий эту модель. В [Goldlucke & Magnor, 2003] для видеопоследовательности осуществляется восстановление трехмерной структуры сцены. В [Cai, 2008] осуществляется построение для всей видеопоследовательности единого изображения, содержащего фон всех кадров видеопоследовательности. В этих методах предполагается, что во всех кадрах рассматриваемой видеопоследовательности площадь объектов переднего плана незначительна по сравнению с площадью фона. Это предположение необходимо для осуществления первоначальной сегментации кадра на область фона и объекты переднего плана. Если маска объектов переднего плана известна (например, она получена путем выделения объектов на каждом кадре вручную), то этап сегментации можно опустить, тем самым освободившись от ограничения на площадь фона. При этом требуется разделение видео на два слоя (фон и объекты переднего плана), что увеличивает вычислительную сложность при обработке видео, содержащего несколько отличающихся по глубине и характеру движения планов.

3 Предложенный метод

Предложенный метод использует карты глубины для каждого кадра исходной видеопоследовательности для автоматического построения маски областей, требующих восстановленного фона при построении ракурсов стерео или многокурсного видео. Это позволяет значительно снизить вычислительную сложность алгоритма, избежав восстановления областей фона, которые затем не будут использоваться. Использование карты глубины также позволяет одновременно восстанавливать фон для нескольких планов исходной видеопоследовательности, поскольку для каждого объекта известно его положение в сцене.

Процесс восстановления фона для каждого кадра видеопоследовательности состоит из некоторой последовательности восстановления фона для пары кадров. Разделы 3.1-3.4 описывают процесс восстановления фона для некоторой пары кадров исходного видео (или процесс попарного восстановления). В разделе 4 описана последовательность, в которой применяются попарные восстановления для достижения наилучшего результата.

Обозначим кадр, для которого происходит восстановление, F_{tar} , а кадр, который является источником для восстановления – F_{src} . Попарное восстановление происходит в 4 этапа:

- 3) Определение для кадров F_{tar} и F_{src} маски восстановления. Разбиение маски восстановления для кадра F_{tar} на блоки для восстановления.
- 4) Анализ движения между кадрами F_{tar} и F_{src} методом блочной оценки движения. Определение для каждого блока восстановления множества векторов кандидатов, определяющих движение фона в рамках данного блока.
- 5) Выбор лучшего вектора для каждого блока восстановления по метрике оценки качества, вычисляемой методом, аналогичным методу сопоставления заплаток.
- 6) Встраивание области фона из кадра F_{src} в кадр F_{tar} по найденному вектору.

Рассмотрим каждый из шагов подробнее.

3.1 Построение маски восстановления

Для определения областей восстановления в кадре используется только соответствующая ему карта глубины. Метод подразумевает, что значение интенсивности пикселя на карте глубины тем больше, чем ближе к наблюдателю расположен объект, содержащий этот пиксель. Поскольку разрывы при сдвиге между пикселями образуются на границах карты глубины, то области, требующие восстановленного фона для построения ракурсов, расположены рядом с границами карты глубины. Таким образом, области карты глубины со значением градиента выше заданного порога (порог определяется динамически, исходя из максимально возможного сдвига пикселей по горизонтали) формируют маску областей восстановления.

После построения маски области восстановления производится ее разбиение на небольшие блоки. Это позволяет аппроксимировать движение каждого блока линейным сдвигом даже в случае сложного характера движения фона. Возможны различные способы этого разбиения, которые влияют на скорость обработки, а также на качество итогового восстановления. Поскольку при построении ракурсов осуществляется сдвиг объектов по горизонтали, то области восстановления, полученные описанным способом, будут иметь вид тонких вертикальных полос вдоль границ объектов. Из-за их небольшой ширины полосы не имеют смысла разбивать их по вертикали, поэтому в описываемом методе используется разбиение на горизонтальные полосы фиксированной высоты (16 пикселей). После разбиения на горизонтальные полосы для каждой полосы определяется, покрывает ли она только край объекта (граница по глубине находится только с одной стороны от нее) или весь объект целиком (границы по глубине расположены с обеих сторон полосы). В случае если полученная полоса покрывает объект полностью (то есть некоторый объект переднего плана полностью попадает внутрь маски области восстановления), измеряется ширина объекта и, если она больше заданного

порога, то полоса разрезается посередине на два отдельных блока восстановления. Дополнительный этап подразбиения подобных блоков необходим в силу используемой стратегии восстановления, описанной в разделе 4. Ограничение разбиения для блоков небольшой ширины введено с целью оптимизации метода. Отдельно стоит отметить, что блоки по построению имеют фиксированную высоту, но произвольную форму. Для каждого блока вычисляется его средняя глубина d_{fg} . В дальнейшем значение d_{fg} используется для корректного определения возможности восстановления значениями интенсивности из другого кадра.

3.2 Анализ движения между кадрами

Анализ движения осуществляется по методу блочной оценки движения, описанном в [Симонян и др., 2008]. Для уточнения получаемого поля векторов движения применяется алгоритм фильтрации векторов движения, описанный в [Зачесов и др., 2012]. В качестве источников векторов-кандидатов используется как исходное поле векторов, так и поле векторов, полученное в результате фильтрации векторов, чтобы исключить влияние возможных ошибок фильтрации на границах объектов.

Для каждого блока восстановления формируется множество векторов-кандидатов, отражающих возможное движение фона, закрытого данным блоком. Изначальное множество кандидатов формируется из всех различных векторов, содержащихся в окрестности текущего блока восстановления, кроме векторов, соответствующие блоки (фиксированной формы и размера, используемые алгоритмом оценки движения) которых полностью или частично лежат внутри маски восстановления, или средняя глубина которых больше d_{fg} текущего блока восстановления.

3.3 Выбор вектора движения блока восстановления

Для выбора лучшего вектора восстановления используется метод сопоставления областей. Для каждого блока восстановления формируется его расширение, которое содержит в себе область кадра F_{tar} , не принадлежащую маске областей восстановления кадра F_{tar} и содержащую пиксели, глубина которых меньше значения d_{fg} . Назовем эту область расширенным блоком восстановления b_{ext} . Для каждого вектора $vcand$ из множества кандидатов осуществляется построение проекции расширенного блока восстановления b_{ext} из кадра F_{src} . Из кадра F_{src} выбираются пиксели, координаты которых (x, y) удовлетворяют $(x, y) = (x_{tar} + vcand_x, y_{tar} + vcand_y)$

где x_{tar}, y_{tar} – координаты пикселя в кадре F_{tar} , а $vcand$ – рассматриваемый вектор-кандидат. В случае вещественных значений координат, для вычисления значения интенсивности пикселя проекции используется билинейная интерполяция между значениями интенсивности

прилегающих пикселей. В случае если направление и величина $vcand$ совпадает с направлением и величиной настоящего вектора движения фона в сцене, то, по построению, полученная проекция расширенного блока будет содержать аппроксимацию той же области фона, что и область b_{ext} кадра F_{tar} , и значения интенсивностей соответствующих пикселей блока b_{ext} и его проекции будут слабо отличаться. Напротив, если $vcand$ сильно отличается от настоящего вектора, то значения соответствующих пикселей будут значительно отличаться. Это условие не будет выполняться только в случае, если значения интенсивности всех пикселей фона в блоке b_{ext} одинаковы, но в этом случае восстановленный фон будет корректным независимо от того, какой вектор-кандидат будет выбран.

Таким образом, для оценки качества вектора-кандидата можно использовать значение среднеквадратичного отклонения между областью b_{ext} кадра F_{tar} и ее проекцией из кадра F_{src} :

$$error = \frac{1}{n} \sum_{b_{ext}} (I(x_{tar}, y_{tar}) - I(x_{proj}, y_{proj}))^2$$

При вычислении значения отклонения используются только пиксели, лежащие вне области восстановления и глубина d которых меньше, чем d_{fg} . Далее это значение называется ошибкой вектора-кандидата.

После получения значений ошибки для каждого вектора-кандидата среди них выбирается вектор с наименьшим значением ошибки. В случае если полученное значение выше порога допустимой ошибки, блок восстановления помечается, как невозстанавливаемый, и для него последний шаг не выполняется. Порог допустимой ошибки является параметром метода и определяет визуальное качество восстановленного фона, но также влияет на площадь восстановленного в результате работы алгоритма фона (чем ниже порог, тем меньше площадь). Порог введен для исключения из рассмотрения областей с резко изменившимся относительно предыдущего кадра фоном, что, как правило, происходит при появлении на фоне движущегося объекта.

3.4 Заполнение блока восстановления

В случае подходящего значения ошибки вектора осуществляется заполнение блока восстановления значениями интенсивности пикселей из проекции блока из кадра F_{src} . При этом используется только пиксели, имеющие в F_{src} глубину $d < d_{fg}$, а также не принадлежащие маске восстановления кадра F_{src} . При переносе значений используется бикубическая интерполяция. Полученные в результате интерполяции значения интенсивности заменяют соответствующие значения пикселей из области восстановления, и пиксель помечается, как восстановленный.

Аналогичная процедура применяется к карте глубины. При построении проекции глубины используется тот же вектор, что и для

построения проекции исходного кадра. Результатом является карта глубины, соответствующая кадру с восстановленным фоном.

4 Стратегии восстановления последовательности кадров

Процесс восстановления фона для каждого кадра последовательности из нескольких (более 2-х) кадров разбивается на попарные восстановления, при каждом из которых используется описанный в разделе 3 метод. Возможны различные стратегии формирования последовательности пар, для которых будет вызываться алгоритм попарного восстановления. Предложенный метод использует стратегию, основанную на распространении во времени областей, для которых восстановленный фон уже известен. Пусть все кадры последовательности пронумерованы от 0 до n . Кадр с номером i обозначим F^i . Попарные восстановления вызываются в следующем порядке:

- Для кадра F^0 восстановление не выполняется
- Для кадра F^1 в качестве исходного кадра для восстановления F_{src} используется кадр F^0 . Восстановленный кадр обозначим F_{tar}^1
- Для кадра F^2 в качестве исходного кадра для восстановления F_{src} используется кадр F_{tar}^1 . Восстановленный кадр обозначим F_{tar}^2
- ...
- Для кадра F^n в качестве исходного кадра для восстановления F_{src} используется кадр F_{tar}^{n-1} . Восстановленный кадр обозначим F_{tar}^n , его же обозначим F_{tar}^n
- Для кадра F_{tar}^{n-1} в качестве исходного кадра для восстановления F_{src} используется кадр F_{tar}^n . Восстановленный кадр обозначим F_{tar}^{n-1}
- Для кадра F_{tar}^{n-2} в качестве исходного кадра для восстановления F_{src} используется кадр F_{tar}^{n-1} . Восстановленный кадр обозначим F_{tar}^{n-2}
- ...
- Для кадра F^0 в качестве исходного кадра для восстановления F_{src} используется кадр F_{tar}^1 . Восстановленный кадр обозначим F_{tar}^0

Полученная последовательность кадров F_{tar}^i является результатом работы стратегии.

Сложность этой стратегии составляет $O(n)$, где n – число кадров в исходном видео. Дополнительным преимуществом является то, что используется весь доступный в видео фон – в восстановлении кадра F^0 могут участвовать пиксели, значения которых были распространены через всю последовательность кадров от кадра n .

5 Результаты

Предложенный алгоритм позволяет восстанавливать фон для произвольных видеопоследовательностей, не делая никаких

предположений относительно характера движения камеры, фона, объектов или площади, занимаемой фоном. На Рис. 1 приведены результаты алгоритма для последовательности со сложным характером движения и несколькими уровнями фона. На Рис. 2 приведены маска восстановления, Исходная карта глубины и восстановленная карта глубины для одного из кадров.



Рис. 3. Вверху: кадры 0 и 30 исходной видеопоследовательности. Внизу: кадры 0 и 30 видеопоследовательности, полученной в результате восстановления фона из движения предложенным методом



Рис. 2. Маска областей восстановления (вверху справа), исходная карта глубины (внизу слева) и восстановленная карта глубины (внизу справа) для 0-го кадра видеопоследовательности

Поскольку поиск наилучшего вектора для каждого блока независим от результатов поиска для остальных блоков, выбор векторов можно осуществлять параллельно. Это позволяет достигнуть значительной производительности метода даже при большом количестве блоков восстановления. Время вычисления восстановленного фона для одной пары кадров приведено в таблице 1.

Таблица 2. Время обработки видеопоследовательностей различного разрешения

Разрешение	Время обработки на Intel Core i7 2.6 ГГц, 8 потоков
720x480	1,5 секунды на кадр
1280x720	3 секунды на кадр
1920x1080	8 секунд на кадр

6 Дальнейшая работа

Основным недостатком описанного метода восстановления является неустойчивость к нелинейным искажениям фона. Несмотря на преимущества стратегии распространения фона в случае, если траектория движения фона или камеры плохо аппроксимируется кусочно-линейной функцией (иными словами, при движении происходит значительный поворот или изменение масштаба изображения), качество и количество

восстановленного фона значительно падают. Причиной этого является заложенное в основу метода использование линейного сдвига для приближения траектории движения фона во втором этапе попарного алгоритма восстановления при достаточно большой площади блока восстановления. Для устранения этого недостатка можно использовать динамическое разбиение блоков восстановления на более мелкие в случае, если суммарная ошибка восстановления подблоков ниже, чем ошибка исходного, большого блока.

Также выбранная стратегия не позволяет корректно обрабатывать ситуации с медленным движением или длинными статичными сценами. Причиной этого является то, что значения пикселей фона всегда берутся из предыдущего кадра. При отсутствии движения одна и та же область фона будет многократно перенесена между кадрами, что приведет к ее размытию из-за повторно применяющейся при каждом переносе интерполяции. С этим можно бороться, применяя гибридную стратегию распространения, которая пропускает кадры или области кадров, в которых движение незначительно (то есть для восстановления кадра F^k берется кадр F^{k-l} , где $l > 1$).

7 Заключение

Описан метод восстановления фона из движения, использующий карту глубины для поиска областей восстановления и анализа их смещения между кадрами. Полученный данным методом восстановленный фон можно использовать при построении стерео или многоракурсного видео для достижения естественного восприятия трехмерной сцены зрителем.

Новизна состоит в значительном уменьшении вычислительной сложности метода за счет использования быстрого алгоритма оценки движения для получения начального приближения – множества векторов-кандидатов движения для каждой области фона. Также алгоритм использует карту глубины для одновременного восстановления фона для нескольких уровней фона и автоматически выполняет построение глубины фона, которая может напрямую использоваться при построении стерео или многоракурсного видео.

8 Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 15-01-08632 а.

Список литературы

- [Воронин и др., 2012]** Воронин В. В., Марчук В. И., Франц В. А., Гапон Н. В. Реконструкция видеопоследовательностей с использованием модели фона // В ж.: Нелинейный мир / Ред. Потапов А.А. Москва: Радиотехника, 2012. 777–781.
- [Зачесов и др., 2012]** Зачёсов А. А., Колинченко М. Б., Ватолин Д. С. Построение карты глубины и многоакурсного видео по стереопаре // В ж.: Мир Техники Кино / Ред. Индлин Ю. А. Москва, 2012, С. 19-21.
- [Симонян и др., 2008]** Симонян К., Гришин С., Ватолин Д.. Адаптивный метод оценки движения в видео. Сборник статей молодых ученых факультета ВМиК МГУ. М. 2008.
- [Cai, 2008]** Xiaodong Cai. Robust Online Video Background Reconstruction Using Optical Flow and Pixel Intensity Distribution. In proceedings of IEEE International Conference on Communications, 2008, pp.526–530.
- [Crivelli et al., 2008]** Crivelli T., Piriou G., Bouthemy P., Cernuschi-Frías B., Yao J. Simultaneous Motion Detection and Background Reconstruction with a Mixed-State Conditional Markov Random Field. In Proceedings of European Conference on Computer Vision, 2008, pp. 113–126.
- [Goldlucke & Magnor, 2003]** Goldlucke B., Magnor M.A.. Joint 3D-reconstruction and background separation in multiple views using graph cuts. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2003, vol. 1, pp. I-683–I-688.
- [Piccardi 2004]** M. Piccardi. Background subtraction techniques: a review. In proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2004, pp. 3099–3104.
- [Sheikh et al., 2009]** Sheikh Y., Javed O., Kanade T. Background Subtraction for Freely Moving Cameras. In proceedings of the IEEE 12th International Conference on Computer Vision, 2009, pp. 1219–1225.