

ВОССТАНОВЛЕНИЕ ТРЁХМЕРНОЙ ГЕОМЕТРИИ СЦЕНЫ ПО ВИДЕОПОТОКУ С ДВИЖУЩЕЙСЯ КАМЕРЫ

Воробей А. А., Недзьведь А. М.

*Белорусский государственный университет, ФПМИ, Минск, Беларусь,
e-mail: captainhorsepower228@gmail.com.*

В последнее время системы дополненной реальности все больше занимают рынок и приобретают популярность. Для их эффективной работы необходимо иметь четкое представление сцены в трехмерном пространстве. В данном случае сцена представляется в виде 3D модели, как совокупность информации о форме и облике объектов. Способы представления 3D сцены включают электронную 3D модель, анаглиф, комбинация исходного изображения и соответствующей ему карты глубины.

Картой глубины (depth map) изображения называют соответствующее ему изображение, в котором каждый пиксель вместо информации о цвете содержит информацию о своей глубине (расстоянии до камеры). Построение такой карты с монокамеры на основе одного видео потока является сложной задачей.

Для реконструкции сцены будет использован видеопоток с движущейся камеры.

1. По двум разнесённым в пространстве кадрам необходимо построить карту соответствий, аналогично карте смещений в стерео-зрении.

2. При помощи триангуляции, карта соответствий будет преобразована в разреженную карту глубин.

3. Опционально, разреженная карта глубин алгоритмически, либо при помощи нейронной сети, может быть преобразована в плотную, гладкую относительно исходного изображения, карту глубины. На этом этапе реконструкция считается завершённой.

Для построения карты соответствий, необходимо произвести ректификацию входных кадров. Тогда эпиполярные линии изображений будут параллельны, а соответствующие линии будут иметь одну и ту же вертикальную координату.

Традиционные методы ректификации, например планарная ректификация, были разработаны для стерео системы с двумя камерами, расположенными вдоль плоскости сцены (fronto parallel). То есть накладывают сильное ограничение на способы движения камеры, а потому не подходят для решения поставленной задачи. Поэтому будет использован алгоритм полярной ректификации, способный ректифицировать изображения даже при движении камеры вдоль главной оптической оси – в случае, когда эпиполярные линии находятся внутри изображения [2].

Суть метода полярной ректификации заключается в переходе к полярным координатам с центром в эпиполюсе. Исходное изображение в качестве эпиполярных линий выступают угловые секторы, ширина которых выбирается так, чтобы на дальнем конце изображения сектор занимал один пиксель. Длина эпиполярных линий сохраняется.

Современные алгоритмы стерео-сопоставления для повышения точности и скорости работы используют свойство локальной гладкости решений полученного решения. Впервые эта идея была реализована в рандомизированном PatchMatch алгоритме [3]. То есть для вычисления смещения очередной части изображения может быть использован результат с соседних, уже обработанных частей. Для

дополнительного повышения скорости может быть построен параллелизм, например на графическом ускорителе, при этом скорость работы при этом может достигать 1000 кадров в секунду на мощной видеокарте, например - Nvidia Titan X.

Однако ввиду дефицита мощных графических ускорителей для персональных компьютеров и полного их отсутствия на мобильных устройствах, используются алгоритм на основе триангуляции. После ректификации изображений, соответствия между ними гарантировано должны находиться вдоль эпиполярных линий, что позволяет значительно сузить область поиска. Из-за вычислительных ошибок при ректификации изображений, а также ошибок определения геометрического положения камеры по отношению к сцене, после фильтрации неверных совпадений, в общем случае, будет получена разреженная карта соответствий (рис. 1).

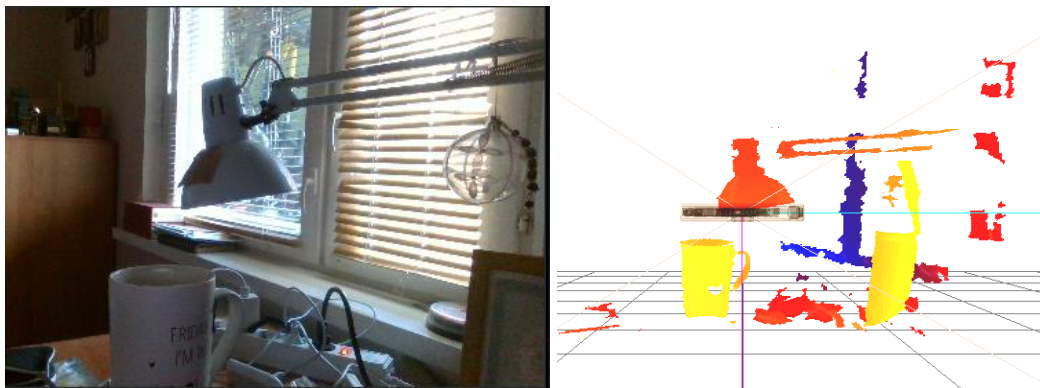


Рис. 1. Разреженная карта соответствий, полученная с камеры ноутбука состоит из 16 уровней глубины и не может захватить все объекты сцены

Задача триангуляции заключается в том, что для некоторой точки трёхмерного пространства X , по двум её проекциями на изображения u, u' необходимо восстановить её трёхмерную координату. Задача сводится к нахождению пересечения двух лучей, тем не менее, всегда присутствует некоторая ошибка геометрической проекции трёхмерного пространства на плоскость, а скрещивающиеся лучи на плоскости не пересекаются на реальной сцене. Поэтому необходимо найти наилучшее решение (приближение), при условии некоторой модели ошибки.

Если ошибка подчиняется стандартному распределению, то наилучшим решением является широко известный в литературе метод оптимальной триангуляции [5]. Этот метод подразумевает поиск минимума полинома шестой степени.

Дополнительно, для интерполяции карты глубины используется билатеральный интерполятор (Fast Bilateral Solver) [6], который решает задачу оптимизации, принимает на вход три изображения: справочное изображение R , целевое изображение T , весовое изображение C (отражает уверенность в данных T). В качестве R может быть использовано исходное изображение с камеры, в качестве T – разреженная карта глубины, в качестве C – маска, обратная к использованной для фильтрации выбросов. В результате получается изображение X , основанное на T , гладкое по отношению к краям объектов на R .

Триангуляция позволяет использовать карту глубины с низким пространственным разрешении, которое свойственно для мобильных устройств, для позиционирования на реальной трёхмерной сцене в задачах дополненной реальности.

Итоговый алгоритм реконструкции 3D сцены состоит из следующей последовательности шагов.

Шаг 1, мобильное устройство записывает видеопоток, при этом для каждого кадра определяются положение в пространстве при помощи доступного на платформе AR-фреймворка.

Шаг 2, для каждого кадра, среди уже известных выбирается оптимальный для поиска стереосоответствий; критерии выбора кадра включают расстояние между кадрами в пространстве, коэффициент пересечения изображений, оценочная ошибка относительной позы кадров.

Шаг 3, для пары кадров проводится полярная стереоректификация и поиск стереосоответствий.

Шаг 4, полученная карта соответствий при помощи триангуляции преобразуется в разреженную карту глубины, которая будет интерполирована до плотной при помощи исходных изображений.

Работа выполнялась при поддержке проекта ГКНТ-Китай Ф20КИТГ-006.

Литература

1. Depth from Motion for Smartphone AR (2018) / ACM Trans. Graph. Vol36, No. 6, Article 193 / J. Valentin, A. Kowdle - с. 3-15
2. A simple and efficient rectification method for general motion / Computer Vision - ECCV'99/ M. Pollefeys, R. Koch - с. 2-5.
3. Многовидовая геометрия в компьютерном зрении / University of Oxford (2000) / А. Зиссерман - с. 250 - 260.
4. Triangulation / Computer Vision and Image Understanding (1997) / R. I. Hartley - с. 146-157.
5. The Fast Bilateral Solver / CVPR (2016) / J. Barron, B. Poole, с 3-12
6. Алгоритмы вычисления оптического потока в задаче определения собственного движения. / Московский Физическо-Технический Институт / Пономарьев Е.С., Григорьев А.С. - с. 353
7. Efficient belief propagation for early vision / CVPR (2004) / Felzenszwalb, P. F. and Huttenlocher, D. P - с. 261-268
8. On the consistency of instantaneous rigid motion estimation / International Journal of Computer Vision (2002) / T. Zhang, C. Tomasi - с. 51-79