

СОВРЕМЕННЫЕ КОНЦЕПЦИИ ВОССТАНОВЛЕНИЯ ТРЕХМЕРНЫХ СЦЕН ПО НАБОРУ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ: НАПОЛНЕНИЕ СИСТЕМ ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ

Юрин Д.В.

*Институт физико-технической информатики, г. Москва, Протвино,
Московский государственный университет им. М.В.Ломоносова, г. Москва,
тел.: 8 (495) 939-11-29, 8 (926) 175-61-28, e-mail: yurin_d@inbox.ru*

В последнее десятилетие наблюдался существенный прогресс в науке о восстановлении трехмерных сцен по двумерным цифровым изображениям. Эти методы весьма востребованы в связи с необходимостью наполнения систем виртуальной реальности данными, соответствующими реальным сценам, в области строительства и архитектуры, зрения мобильных роботов. В настоящее время интерес исследователей смещается в область восстановления трехмерных сцен по одному изображению, что хорошо применимо для искусственных сцен, в основном состоящих из прямых линий и плоскостей [1,2]. Однако, в рамках настоящего обзора, сосредоточимся на алгоритмах восстановления по стереопаре (Standard and Wide-base Stereo), в особенности на стерео с широкой базой съемки и восстановлении по набору изображений (>2, Shape from Motion) из-за применимости этих подходов для обработки широкого класса сцен, наивысшей точности, возможности оценки достоверности результата и существенной роли этих методов в других классах алгоритмов восстановления трехмерных сцен. Отметим здесь, что алгоритмы стандартного стерео предполагают малую по сравнению с глубиной сцены базу съемки и ректифицированные (rectified) изображения. Последнее означает, что эпиполярные линии горизонтальны и находятся в строках с одинаковым номером обоих изображений. Это соответствует ситуации строго параллельных оптических осей камер, если это условие не вполне выполняется, то возможна алгоритмическая ректификация [3], что тесно связано с методами рассматриваемыми ниже. Для алгоритмов стандартного стерео в настоящее время разработана единая методика тестирования, сравнения и автоматического рейтинга стерео алгоритмов [4], в авторитетных научных изданиях новые алгоритмы, не прошедшие этой процедуры не рассматриваются. Большой прогресс достигнут с 1999 года в алгоритмах основанных на графах [5], особенно интересно сочетание стандартного стерео с сегментацией [6,7].

Класс алгоритмов восстановления 3D по стерео и множеству изображений, по сути, основан на сопоставлении точек объекта видимых на разных изображениях и на возможности восстановления высоты треугольника по основанию и двум прилежащим углам. Значительный прогресс в этой области обусловлен 1) глубоким развитием методов проективной геометрии применительно к области компьютерного зрения [1] и 2) революционными достижениями в области детектирования характеристических точек на изображениях и сопоставлением их между изображениями [8,9,10]. Второе обусловлено разработкой теории пространства переменных разрешений [11], теорий цветовых [12] и дифференциальных [13,14] инвариантов и созданием дескрипторов характеристических точек класса SIFT, SURF, GLOH [8,9,10]. Список литературы с комментариями по наиболее важным достижениям в области трехмерного зрения можно найти в [15].

Кратко о геометрии зрения. Считается, что камера проективная, т.е. полагая, что начало системы координат совпадает с оптическим центром камеры, координаты точки на изображении (x,y) связаны с трехмерными координатами точки сцены (X,Y,Z) соотношением

нием $\frac{x}{f} = \frac{X}{Z}$, $\frac{y}{f} = \frac{Y}{Z}$, где f – фокусное расстояние, то есть образом любой прямой линии

в трехмерном пространстве будет прямая линия на изображении. Это требование выполняется приближенно из-за наличия радиальной дисторсии (бочка) у реальных объективов. Следует иметь в виду, что задача восстановления эпиполярной геометрии весьма чувствительна к подобным искажениям, поэтому коррекция радиальной дисторсии *строго обяза-*

тельно, что обычно осуществляется путем предварительной калибровки камеры [16,17], считая, что за время между калибровкой и реальной съемкой калибровочные характеристики не изменились. В случае, когда камера не доступна для проведения калибровки, можно воспользоваться поиском предположительно прямых линий на изображениях и найти преобразование, распрямляющие такие линии [17,18].

Базовыми понятиями проективной геометрии и трехмерного зрения являются гомография, эпиполярная геометрия, связанная с фундаментальной (или существенной) матрицей (ФМ) и трифокальный тензор. Если две камеры наблюдают один и тот же плоский объект в трехмерном пространстве, то однородные координаты [1,16] $\bar{x}^T = (x, y, 1)$ и $\bar{x}'^T = (x', y', 1)$ соответствующих точек этой плоскости на изображениях связаны соотношением $\bar{x}' = \mathbf{H}\bar{x}$, где \mathbf{H} называется матрицей гомографии и определена с точностью до множителя, т.е. имеет 8 степеней свободы. Какая бы ни была трехмерная сцена, не деформирующаяся между получением двух изображений (rigid body), образы трехмерных точек на изображениях связаны соотношением $\bar{x}'^T \mathbf{F} \bar{x} = 0$, матрица \mathbf{F} : $rank \mathbf{F} = 2 \Rightarrow \det \mathbf{F} = 0$ называется фундаментальной (ФМ), она тоже определена с точностью до множителя, и, таким образом, имеет 7 степеней свободы. Матрица \mathbf{F} определяется только взаимным расположением камер и их внутренними параметрами и не зависит от трехмерной сцены. Легко видеть, что для каждой точки \bar{x} на одном изображении, это соотношение верно для всех точек \bar{x}' второго изображения, лежащих на прямой линии $\bar{l} = \mathbf{F}^T \bar{x}$, называемой эпиполярной. Такие ограничения называются эпиполярными, ключевым моментом трехмерного зрения является контроль этих ограничений: для недеформируемой сцены, их нарушение является физически невозможным, и алгоритмы не должны рассматривать невозможные ситуации. На каждом изображении эпиполярные линии пересекаются в одной точке, называемой эпиполлюсом. Если на изображениях найдены прямые линии (концевые точки не определены), то для двух изображений не существует подобных геометрических ограничений, они есть для трех изображений и задаются через трифокальный тензор [1].

С современной точки зрения, восстановление трехмерной сцены **должно состоять из следующих шагов**.

1) Детектирование на изображениях характеристических точек. Детектор обязательно должен выполнять три шага: а) детектирование точек на каждом уровне детальности Scale Space, определение масштаба, где особенность проявляется наилучшим образом, б) уточнение масштаба и пространственных координат точки [28], аффинная нормализация окрестности с дополнительным уточнением координат желательна в) построение дескриптора (вектора параметров), который должен быть не хуже, чем у SIFT, SURF, GLOH, Color SIFT. В этом случае наложение ограничений типа малости различий между кадрами, как было в ранних детекторах не обязательно, зум также может различаться значительно, например в [19] зум различался до 6 раз. Хорошей идеей является коррекция бочки непосредственно преобразованием координат точек найденных на исходных изображениях, чтобы избежать влияния ресамплинга изображений на точность детектирования точек.

2) Первичное установление соответствий между точками путем поиска точки, минимально отличающейся по вектору параметров. Следует оставить только надежные соответствия, такие, что расстояние до следующей точки заметно больше найденного (в $\sim 1.5-2$ раза). Желательно найти 100-400 соответствий, но не менее 30-50.

3) Поиск ФМ с помощью алгоритма 11.4 из [1]. Обязательны все три этапа: 7-точечный алгоритм, погруженный в RANSAC, 8-точечный алгоритм по всем инлайерам, нелинейная минимизация. Роль процедуры RANSAC состоит в отсеиве не только ложных соответствий, но и вырожденных конфигураций (эти 7 точек в 3D лежат на поверхности 2-го порядка) или конфигураций, приводящих к низкой точности. Возможные варианты: первоначально искать Gold Standard алгоритмом гомографии [1], это исключает широкий класс вырожденных ситуаций или своевременно о них сигнализирует. Если гомография найдена, ее можно уточнить, применяя ко всей такой области алгоритмы оптического по-

тока [20]. Если найдена только одна плоскость, занимающая значительную часть изображений, вместо 7-точечного алгоритма можно использовать 6-точечный, что благотворно влияет на RANSAC. Если найдено 2 плоскости, то из 2 гомографий можно найти ФМ [21]. Если первичных точечных соответствий было найдено недостаточно, существует алгоритм [22] поиска ФМ по прямым линиям, о поиске и соотнесении таких линий см. [27].

4) Имея эпполярные ограничения (ФМ), установить соответствия между незадействованными характеристическими точками, сопоставляя по вектору параметров только точки, лежащие в коридоре погрешности [23] вдоль эпполярной линии. Уточнить ФМ.

5) По ФМ можно восстановить 3D координаты точек в единицах базы и ориентацию камер [1,24]. В случае если кадров больше 2, следует путем анализа попарных соответствий найти точки, для которых соответствия удастся проследить по большему количеству кадров. Если удалось найти хотя бы 10-20 точек прослеженных по всей последовательности кадров (>4), метод [25] позволяет провести более устойчивое и самосогласованное восстановление координат точек, камер и ориентаций с одновременной оценкой погрешности результата. Если таких точек недостаточно, или кадров менее 4, то можно воспользоваться тем, что если вычислены ФМ между всеми парами кадров, то эпполярные линии, соответствующие разным начальным кадрам, на общем конечном, в большинстве случаев, будут пересекаться. Это приводит от ограничения «искать вдоль прямой» к ограничению «искать в окрестности пересечения прямых», что дает возможность проверить достоверность установленных соответствий и отсеять ошибочные соответствия.

6) Найденные на шаге 1 характеристические точки, как правило, лежат на локально плоских участках, а не на гранях сцены, поэтому построение сеточной модели по такому облаку 3D точек дает неадекватные результаты. Пользуясь найденными эпполярными ограничениями, следует добавить виртуальные точки, образованные пересечением линий (edge, см. [15]), найденных на изображениях, с эпполярными линиями. В качестве дескриптора для таких точек можно использовать подход из [27]. Нахождение 3D координат таких точек тривиально, поскольку расположение и ориентация камер уже известны.

7) Полученное облако 3D точек следует проанализировать на предмет не выявленных ранее плоскостей, область ими занимаемая может быть расширена путем репроецирования и анализа разностного кадра, а положения плоскостей уточнены [26].

8) Для неплоских участков трехмерной сцены – построение поверхности путем триангуляции Делоне. Желателен поиск гладких аппроксимаций, этот вопрос в настоящее время недостаточно изучен, однако хорошей начальной точкой может послужить адаптация подходов [6,7] в применении к *фрагментам* сцены.

9) Уточнение границ путем решения задачи сегментации (разметки) на графах. Найден ряд поверхностей, для каждого пикселя каждого кадра найти метку, обозначающую, к какой поверхности его следует отнести, за отнесение соседних точек к различным поверхностям - штраф. Здесь следует отметить, что при восстановленной сцене и расположении камер, для каждого пикселя одного кадра выбор поверхности однозначно устанавливает соответствующие ему пиксели всех остальных кадров. Их непохожесть между собой формирует невязку.

Работа выполнена при финансовой поддержке РФФИ, гранты: 08-07-00468-а, 08-07-00469-а, 09-01-00703-а, 09-01-92470-МНКС_а.

Литература

1. Hartley R., Zisserman A. Multiple View Geometry in Computer Vision. – Cambridge University Press, 2004. – 672 p., ISBN: 0521540518.
2. Make3D publications, <http://make3d.stanford.edu/publications.html>.
3. Hartley R.I. Theory and Practice of Projective Rectification // International Journal of Computer Vision. – 1999. – V.35. – No. 2. – P. 115 – 127.
4. Единая система тестирования стерео алгоритмов <http://vision.middlebury.edu/sterео/>.
5. Kolmogorov V., Zabih R. Computing visual correspondence with occlusions using graph cuts

- // In International Conference on Computer Vision, Vancouver, Canada, July 2001.
6. *Birchfield S., Tomasi C.* Multiway cut for stereo and motion with slanted surfaces // Proceedings of International Conference on Computer Vision. – 1999, – P. 489 – 495.
 7. *Lin M.H., Tomasi C.* Surfaces with Occlusions from Layered Stereo. //IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003.
 8. *Tuytelaars T., Mikolajczyk K.* **Local Invariant Feature Detectors: A Survey** // Foundations and Trends® in Computer Graphics and Vision. – 2008. – V. 3. – No 3, – P. 177 – 280.
 9. *Mikolajczyk K., Schmid C.* **A performance evaluation of local descriptors** //IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2005. – V. 27, – No. 10, – P.1615 – 1630.
 10. *Burghouts G.J., Geusebroek J.M.* **Performance evaluation of local colour invariants** // Computer Vision and Image Understanding. – 2009. – V. 113, – P. 48 – 62.
 11. *Lindeberg T.* Scale–Space Theory in Computer Vision // Kluwer Academic Publishers, Dordrecht, Netherlands. – 1994. <http://www.nada.kth.se/~tony/earlyvision.html>.
 12. *Geusebroek J.M., van den Boomgaard R., Smeulders A.W.M., Geerts H.* **Color invariance.** IEEE Trans. Pattern Anal. Machine Intell. – 2001. – V. 23. – No. 12. – P. 1338 – 1350.
 13. *Schmid C., Mohr R.* **Local Grayvalue Invariants for Image Retrieval.** //IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). – 1997. – V. 19. – No. 5. – P. 530 – 534.
 14. *Florack L.M.J., ter Haar Romeny B.M., Koenderink J.J., Viergever M.A.* **Cartesian Differential Invariants in Scale–Space** // Journal of Mathematical Imaging and Vision. – 1993, – V. 3, – P. 327 – 348.
 15. Ссылки на важнейшие работы: http://imaging.cs.msu.su/~yurin/basic_literature.html.
 16. *Mohr R., Triggs B.* Projective Geometry for Image Analysis // A Tutorial given at ISPRS, Vienna, July 1996. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.3924>.
 17. *Devernay F., Faugeras O.* Straight lines have to be straight: Automatic calibration and removal of distortion from scenes of structured environments //Machine Vision and Applications. – 2001. – V. 13. – P. 14 – 24. <http://devernay.free.fr/publis/distcalib-mva.pdf>.
 18. *Fitzgibbon A. W.* **Simultaneous Linear Estimation of Multiple View Geometry and Lens Distortion** // Proceedings on CVPR. – 2001.
 19. *Dufournaud Y., Schmid C., Horaud R.* **Matching images with different resolutions** //Proceedings on CVPR, – 2000 (Hilton Head Island, SC, USA). – V. 1, – P. 612 – 618.
 20. *Zokai S., Wolberg G.* Image Registration Using Log–Polar Mappings for Recovery of Large–Scale Similarity and Projective Transformations //IEEE Transactions on Image Processing – 2005. – V. 14, – No. 10. <http://www-cs.engr.ccnycunyu.edu/~wolberg/pub/tip05.pdf>.
 21. *Luong Q.T., Faugeras O.D.* **Determining the fundamental matrix with planes: instability and new algorithms** //Proceedings on CVPR 1993. – P.489 – 494.
 22. *Sagues C., Murillo A.C., Escudero F., Guerrero J.J.* **From lines to epipoles through planes in two views** //Pattern Recognition – 2006. – V. 39. – No. 3. – P. 384 – 393.
 23. *Csurka G., Zeller C., Zhang Z., Faugeras O.D.* **Characterizing the Uncertainty of the Fundamental Matrix** //Computer Vision and Image Understanding. – 1997. – V.68. – No. 1. – P. 18 – 36.
 24. *Longuet–Higgins H.C.* A Computer Algorithm for Reconstructing a Scene From Two Projections // Nature, September 1981. – V. 293, – P. 133 – 135.
 25. *Свешникова Н.В., Юрин Д.В.* **Алгоритмы факторизации: достоверность результата и применение для восстановления эпиполярной геометрии** // Труды конференции Графикон 2006, Россия, Новосибирск, Академгородок, 1 – 5 июля 2006. – С.158 – 165.
 26. *Semeikina E., Yurin D.* **Finding and refinement planes in 3D points cloud obtained under 3D recovery from image set** // Proceedings of GraphiCon'2008, Moscow, Russia, 2008, P.306.
 27. *Волегов Д.Б., Юрин Д.В.* **Предварительное грубое совмещение изображений по найденным на них прямым линиям для построения мозаик, сверхразрешения и восстановления трехмерных сцен** //Программирование – 2008, – Т.34, – N.5, –С.47 – 66.
 28. *Brown M., Lowe D.* **Invariant Features from Interest Point Groups** //In British Machine Vision Conference. – 2002. – P. 656 – 665. <http://citeseer.ist.psu.edu/708522.html>.