

СОВРЕМЕННЫЙ ВЗГЛЯД НА СТРУКТУРУ СИСТЕМ АВТОМАТИЧЕСКОГО ПОСТРОЕНИЯ ТРЕХМЕРНЫХ ВИРТУАЛЬНЫХ МОДЕЛЕЙ ПО ИЗОБРАЖЕНИЯМ (ОБЗОР)

Юрин Д.В. (с.н.с., к.ф.-м.н.)

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

Значительный прогресс в области алгоритмов восстановления 3D по двум и более изображениям [1] обусловлен 1) глубоким развитием методов проективной геометрии применительно к области компьютерного зрения [2] и 2) революционными достижениями в области детектирования характеристических точек на изображениях и сопоставлением их между изображениями [3,4,5], основанными на теории пространства переменных разрешений [6], теориях цветовых [7] и дифференциальных [8,9] инвариантов и созданием дескрипторов таких точек класса SIFT, SURF, GLOH [3,4,5].

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 $x/f = X/Z$, $y/f = Y/Z$, где f – фокусное расстояние, т.е. образом любой прямой линии в трехмерном пространстве будет прямая линия на изображении. Следует иметь в виду, что задача восстановления эпиполярной геометрии (см. далее) весьма чувствительна к радиальной дисторсии (бочке), присутствующей у реальных объективов, поэтому ее коррекция [10,11,12] **строго обязательна**.

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 [2,10] $\vec{x}^T = (x, y, 1)$ и $\vec{x}'^T = (x', y', 1)$ соответствующих точек этой плоскости на изображениях связаны соотношением $\vec{x}' = \mathbf{H}\vec{x}$, 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 $\vec{x}^T \mathbf{F} \vec{x}' = 0$, 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 \vec{x} на одном изображении, это соотношение верно для всех точек \vec{x}' второго изображения, лежащих на прямой линии $\vec{l} = \mathbf{F}^T \vec{x}$, 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 **эпиполусом**. Если на изображениях найдены прямые

линии (концевые точки не определены), то для двух изображений не существует подобных геометрических ограничений, но они есть для трех изображений и задаются через трифокальный тензор [2].

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 [13], желательна аффинная нормализация окрестности с дополнительным уточнением координат; в) построение дескриптора (вектора параметров), который должен быть не хуже, чем у SIFT, SURF, GLOH, Color SIFT [5]. В этом случае наложение ограничений типа малости различий между кадрами, как было в ранних детекторах, не обязательно, зум также может значительно различаться, например в [14] 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 ~1.5-2 раза). Желательно найти 100-400 соответствий, но не менее 30-50.

3) Поиск ФМ с помощью алгоритма 11.4 из [2]. Обязательны все три этапа: 7-точечный алгоритм, погруженный в RANSAC, 8-точечный алгоритм по всем инлайерам, нелинейная минимизация. Роль процедуры RANSAC состоит в отсеивании 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 гомографии Gold Standard алгоритмом 4.3 из [2], это исключает широкий класс вырожденных ситуаций или своевременно о них сигнализирует. Если гомография найдена, ее можно уточнить, применяя ко всей такой области алгоритмы оптического потока [15]. 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 [16]. Если первичных точечных соответствий было найдено недостаточно, существует алгоритм [17] поиска ФМ по прямым линиям, о поиске и соотношении таких линий (дескрипторы, и их сравнение) см. [18].

4) Имея эпиполярные ограничения (ФМ), установить соответствия между незадействованными характеристическими точками, сопоставляя по вектору параметров только точки, лежащие в коридоре погрешности [19] вдоль эпиполярной линии. Уточнить ФМ. Применительно к алгоритмам стандартного стерео, фундаментальная матрица позволяет выполнить ректификацию стереопары [20].

5) По ФМ можно восстановить 3D координаты точек в единицах базы и ориентацию камер [2,21]. 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 [22] позволяет провести более устойчивое и самосогласованное восстановление координат точек, камер и ориента-

ций с одновременной оценкой погрешности результата. Если прослежено недостаточно точек или кадров менее 4, то следует воспользоваться тем, что эпиполярные линии на одном кадре, определяемые остальными камерами, в большинстве случаев, будут пересекаться, Это приводит от ограничения «искать вдоль прямой» к ограничению «искать в окрестности пересечения прямых», что позволяет 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 эпиполярными ограничениями, следует добавить виртуальные точки, образованные пересечением линий (edge, см. [1]), найденных на изображениях, с эпиполярными линиями. В качестве дескриптора для таких точек можно использовать подход из [18]. Так как расположение и ориентация камер уже известны, вычисление 3D координат тривиально.

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 [23].

8) Для неплоских участков трехмерной сцены – построение поверхности путем триангуляции Делоне. Желателен поиск гладких аппроксимаций, хорошей начальной точкой может послужить адаптация подходов [24,25] в применении к *фрагментам* сцены.

9) Уточнение границ путем решения задачи сегментации на графах [24]: 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. При восстановленной сцене и расположении камер, для каждого пикселя одного кадра выбор поверхности однозначно устанавливает соответствующие ему пиксели остальных кадров. Разность их цвета формирует невязку.

Работа выполнена при поддержке РФФИ, гранты: 08-07-00468-а, 08-07-00469-а, 09-01-92470-МНКС_а, 09-07-92000-ННС_а.

Литература

1. Ссылки на важнейшие работы по 3D и др.: http://imaging.cs.msu.ru/~yurin/basic_literature.html.
2. *Hartley R., Zisserman A. Multiple View Geometry in Computer Vision.* – Cambridge University Press, 2004. – 672 p., ISBN: 0521540518.
3. *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.
4. *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.
5. *Burghouts G.J., Geusebroek J.M. Performance evaluation of local colour invariants* // Computer Vision and Image Understanding. – 2009. – V. 113, – P. 48 – 62.
6. *Lindeberg T. Scale–Space Theory in Computer Vision* // Kluwer Academic Publishers, Dordrecht, Netherlands. – 1994. <http://www.nada.kth.se/~tony/earlyvision.html>.
7. *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.
8. *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.
9. 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.
 10. 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>.
 11. 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>.
 12. Fitzgibbon A. W. Simultaneous Linear Estimation of Multiple View Geometry and Lens Distortion // Proceedings on CVPR. – 2001.
 13. 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>.
 14. 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.
 15. 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.cuny.cuny.edu/~wolberg/pub/tip05.pdf>.
 16. Luong Q.T., Faugeras O.D. Determining the fundamental matrix with planes: instability and new algorithms //Proceedings on CVPR 1993. – P.489 – 494.
 17. 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.
 18. 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.
 19. 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.
 20. Hartley R.I. Theory and Practice of Projective Rectification // International Journal of Computer Vision. – 1999. – V.35. – No. 2. – P. 115 – 127.
 21. Longuet–Higgins H.C. A Computer Algorithm for Reconstructing a Scene From Two Projections // Nature, September 1981. – V. 293, – P. 133 – 135.
 22. 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.
 23. 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.
 24. Birchfield S., Tomasi C. Multiway cut for stereo and motion with slanted surfaces // Proceedings of International Conference on Computer Vision. – 1999, – P. 489 – 495.
 25. Lin M.H., Tomasi C. Surfaces with Occlusions from Layered Stereo. //IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003.