

## ВИЗУАЛЬНАЯ НАВИГАЦИЯ С ВРЕМЯПРОЛЕТНОЙ КАМЕРОЙ

© 2010 г. С. В. Серебряков\*, Л. А. Станкевич\*\*, канд. техн. наук

\* Университет Якобс, Бремен, Германия

\*\* Санкт-Петербургский государственный политехнический университет, Санкт-Петербург

E-mail: cfr.ssv@gmail.com, stankevich\_lev@inbox.ru

Представлена система визуальной навигации в реальном времени с использованием времяпролетной камеры без априорных знаний о сцене. Рассмотрен способ комплексирования времяпролетной и оптической камер. Рассмотрены методы повышения надежности локализации, учитывающие цветовую и пространственную информацию.

**Ключевые слова:** времяпролетная камера, компьютерное зрение, робототехника, локализация.

Коды OCIS: 100.2960

Поступила в редакцию 14.05.2010

### Визуальная одновременная локализация и навигация

Задачу составления карты роботом часто называют задачей одновременной локализации и составления карты (*Simultaneous Localization And Mapping*, далее SLAM [1]). Робот не только обязан составить карту, но и должен сделать это, изначально не зная своего местоположения. Проблему можно сформулировать двумя вопросами робота: “Где я?” и “Как выглядит мир?”. Ответить на первый вопрос можно, только зная ответ на второй и наоборот. SLAM – одна из наиболее важных задач робототехники, решение которой ведет к созданию действительно автономных роботов [2]. В качестве основного устройства восприятия часто используется лазерный сканер [3], однако в последнее время все больше возрастает интерес к системам на основе компьютерного зрения, так как такие системы пассивны, на порядок дешевле и компактнее, а поток информации в разы больше (что особенно важно в интерпретации сцен). Самым главным недостатком является срыв слежения и сопоставления на слаботекстурированных или засвеченных сценах (например, в однотонно окрашенных помещениях или при попадании яркого источника света в поле зрения камеры). В робототехническом сообществе глубоко изучена проблема визуальной навигации (vSLAM) с двумя и более камерами и в различных комбинациях с другими датчиками [4–8]. Также

интерес представляют монокулярные системы, активно изучаемые как робототехникой [9], так и компьютерным зрением [10–12]. Помимо робототехнических, приложениями vSLAM также являются дополненная реальность и человеко-машинные интерфейсы.

### Времяпролетные камеры

Недавно благодаря развитию микроэлектроники стали доступны новые устройства для трехмерного восприятия – времяпролетные камеры (ВК, англ. *time-of-flight camera*). ВК-устройство, работающее по схожему с лазерными сканерами принципу, основными частями которого являются излучатель (обычно инфракрасный) и матричный фотоприемник. Для каждого пиксела определяется время, потребовавшееся для путешествия света до объекта. Более подробно принципы работы ВК описаны в работе [13], в статье [14] ВК сравнивается со стереосистемой.

Времяпролетные камеры сочетают плотность реконструкции лазерных сканеров и компактность систем на основе компьютерного зрения. Удачными примерами использования данной технологии являются трехмерная реконструкция [15], захват движения [16], слежение за лицом [17], а также в человеко-машинных интерфейсах.

В данной работе использовалась PMD-камера ОЗ с разрешением 64×48 пикселей и рабочей

дистанцией до 7 метров, соединенная с цветной веб-камерой с разрешением 640×480. Ошибка измерения для каждого пиксела составляет 20–200 мм в зависимости от дистанции.

### SLAM с времяпролетной камерой

**Калибровка.** ВК создает цифровую матрицу дистанций до объектов сцены в каждом пикселе, для вычисления трехмерных координат точек необходимо откалибровать камеру (рис. 1). В соответствии с моделью перспективной проекции [18]

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} [Rt] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (1)$$

где  $\mathbf{A}$  – матрица внутренних параметров камеры

$$\mathbf{A} = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

получаем выражение для трехмерной точки  $(x, y, z)$ , соответствующей пикселу  $(u, v)$  с глубиной  $r$

$$s = \sqrt{c_x 2f_x^2 + c_x^2 f_x^2 + f_x^2 f_y^2 - 2c_x f_y^2 u + f_y^2 u^2 - 2c_y f_x^2 v + f_x^2 v^2}, \quad (3)$$

$$\begin{aligned} x &= f_y r (c_x - u) / s, \\ y &= f_x r (c_y - v) / s, \\ z &= f_x f_y / s. \end{aligned} \quad (4)$$

Для устранения дисторсии используется стандартная полиномиальная модель. Калибровка выполняется с помощью метода Жанга [19] для обеих камер. Уравнения (3), (4) при  $r = 1$  позволяют получить соответствующий луч для каждой точки изображения веб-камеры. Помимо дисторсии измерения ВК содержат различные виды ошибок: случайные (ошибки дискретизации, квантовый шум, рассеяние света) и систематические.

Имея декартовы координаты точек с ВК, можно перепроецировать их на матрицу веб-камеры в соответствии с моделью (1) и найти соответствующие пиксели (рис. 2, 3).

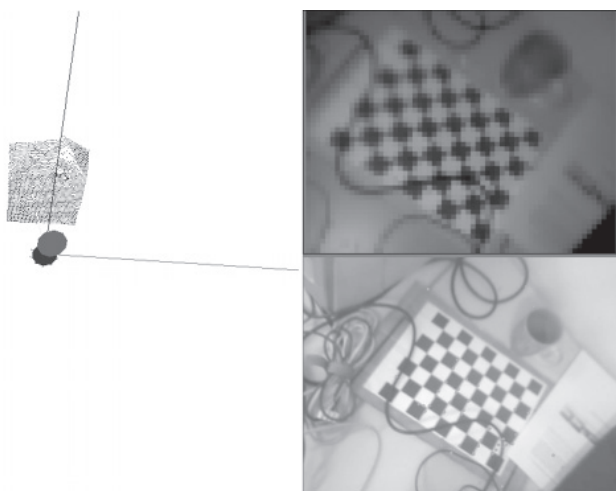


Рис. 1. Калибровка ВК.



Рис. 2. Облако точек с разными дальностями.

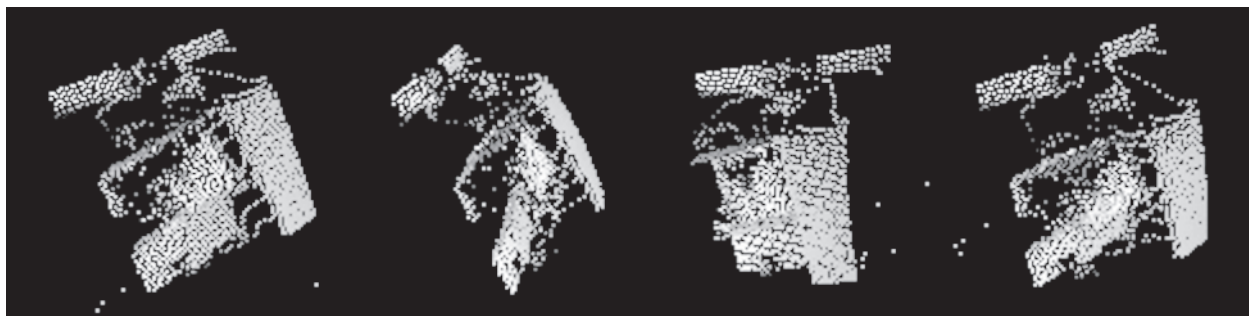


Рис. 3. Фрагмент реконструированного объекта (шкафа).

ритмы “*Good Features To Track*” [21] и FAST (*Features from Accelerated Segment Test*) [22].

Отслеживая движение точек на изображениях с веб-камеры, можно перепроецировать их на ВК (1) и получить соответствия трехмерных точек для последовательных кадров. С их помощью можно просчитать движение между

кадрами и, интегрируя, получить траекторию камеры.

Имея три пары точек  $(l_1, l_2, l_3)$  и  $(r_1, r_2, r_3)$ , можно определить движение  $[Rt]$ , отбросив лишние ограничения (для решения необходимы только 6 уравнений). Наиболее быстрым методом является прямой метод Хорна [23]

$$\begin{aligned}\hat{x}_1 &= (l_2 - l_1) / \|l_2 - l_1\|, \\ \hat{y}_1 &= \left\{ (l_3 - l_1) - [(l_3 - l_1) \circ \hat{x}_1] \hat{x}_1 \right\} / \left\| (l_3 - l_1) - [(l_3 - l_1) \circ \hat{x}_1] \hat{x}_1 \right\|, \\ \hat{z}_1 &= \hat{x}_1 \times \hat{y}_1, \\ M_1 &= [\hat{x}_1 \hat{y}_1 \hat{z}_1], \\ R &= M_2 M_1^T.\end{aligned}\tag{5}$$

После нахождения вращения, перемещение может быть определено вычитанием центроидов

$$\begin{aligned}t &= c_l - c_r, \\ c_l &= (Rl_1 + Rl_2 + Rl_3) / 3, \\ c_r &= (r_1 + r_2 + r_3).\end{aligned}\tag{6}$$

К сожалению, прямое решение очень неустойчиво к выбросам, поэтому были использованы более надежные методы, описанные ниже. Однако благодаря скорости прямое решение может быть использовано в составе алгоритма RANSAC (*Random Sample Consensus*) [24], что позволит добиться быстрого и надежного решения. RANSAC хорошо подходит и для создания алгоритмов в параллельной архитектуре.

**Локализация.** Для определения движения использован алгоритм ICP (*Iterative Closest Point*, итеративный алгоритм близких точек), выравнивающий облака точек. Алгоритм был разработан независимо авторами работ [25–27] и является самым популярным для выравнивания облаков точек. Пусть  $M = \{m_j | j = 1, \dots, N_m\}$  – множе-

ство точек предыдущего кадра (модель), а  $G = \{g_i | i = 1, \dots, N_g\}$  – множество точек текущего кадра (сцена). Тогда для каждой точки сцены определяется ближайшая точка модели

$$m_k = \arg \min_{i=1, \dots, N_m} \|g_k - m_i\|.\tag{7}$$

Алгоритм ICP находит преобразование  $[R|t]$  с помощью итеративной минимизации наименьших квадратов. В общем виде функцию ошибок можно записать как

$$E(R, t) = \sum \sum w_{i,j} \|(Rg_i + t) - m_j\|^2,\tag{8}$$

где  $w_{i,j} = 1$  для соответствующих точек  $(g_i, m_i)$  и нулю для несоответствующих. Существует четыре алгоритма решения этой задачи в закрытой форме, подробный анализ выполнен в работе [28].

Данный алгоритм оказался недостаточно устойчивым для определения движения по небольшому количеству отслеживаемых точек, но приемлемо выравнивал полные облака на сценах с выраженной структурой.



**Увеличение надежности локализации.** Для улучшения детектирования выбросов можно использовать цветовую информацию. Из-за низкого разрешения ВК в местах разрывов глубины сцены возникают “фантомные точки”, поверхности, соединяющие грани разрыва. Так как разрывам в пространстве обычно соответствуют разрывы в цвете пикселей на изображении, можно построить детектор выбросов на основе цветовой информации и распределении глубин [29]. Приняв предположение о гауссовом распределении измерений глубины, вычислим надежность измерения для каждого  $i$ -пикселя

$$s_i = \sum \exp \left[ -w_{ij} (d_i - d_j)^2 / \sigma_d \right]. \quad (9)$$

Суммирование проводится по окружающим пикселям, а вес вычисляется из его цвета

$$w_{ij} = \exp \left( -\|c_i - c_j\| / \sigma_c \right). \quad (10)$$

В результате получаем большую чувствительность детектора на равномерно окрашенных поверхностях, обычно соответствующих цельным поверхностям, надежно измеряемым ВК. Коэффициенты  $\sigma_d$  и  $\sigma_g$  соответствуют характеристикам камеры и настраиваются эмпирически.

### Заключение

Времяпролетная камера позволяет получать облако точек в реальном времени без дополнительных вычислительных затрат. Уже сейчас разрешение и цена делают ВК достойной альтернативой стереосистемам. Рассмотренная система была протестирована на разных сценах. Несмотря на фильтрацию приемлемые результаты были получены только на сценах с выраженной структурой. Из-за небольшого угла зрения и низкого разрешения алгоритм ICP намного хуже справляется с выравниванием, чем при использовании лазерного сканера в качестве основного сенсора.

На реконструированном фрагменте (рис. 4) видно, что лучшее качество достигнуто на детализированных фрагментах (полка), а на более плоских присутствует большое количество ложных кадров. Данные обстоятельства ограничивают использование ВК в системах реконструкции однородных сцен, однако этот сенсор может быть идеальным решением для систем реконструкции небольших или детализированных объектов, обхода препятствий и распознавания образов.

Исследования были выполнены на грант Федерального Министерства Образования и Исследований Германии (BMBF) RUS09/24.

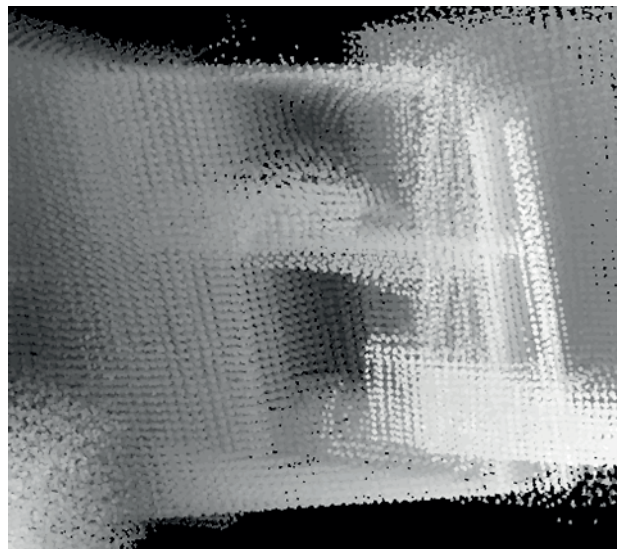


Рис. 4. Пример реконструкции методом ICP.

### ЛИТЕРАТУРА

1. Agrawal M., Konolige K. Real-time localization in outdoor environments using stereo vision and inexpensive GPS // Proc. 18<sup>th</sup> Inter. Conf. on Pattern Recognition. Hong Kong. August 20–24 2006. P. 1063–1068.
2. Beder C., Bartczak B., Koch R. A combined approach for estimating patchlets from PMD depth images and stereo intensity images // Pattern Recognition. 2007. V. 4713. № 2. P. 11–20.
3. Besl P.J., McKay H.D. A method for registration of 3D Shapes // Pattern Analysis and Machine Intelligence. 1992. V. 14. № 2. P. 239–256.
4. Chen Y., Medioni G. Object modeling by registration of multiple range images // Image and Vision Computing. 1991. V. 10. № 3. P. 2724–2729.
5. Davison A.J., Reid I.D., Molton N.D., Stasse O. Monoslam: Real-time single camera SLAM // Pattern Analysis and Machine Intelligence. 2007. V. 29. № 6. P. 1052–1067.
6. Eade E., Drummond T. Edge landmarks in monocular SLAM // Proc. British Machine Vision Conference (BMVC). Edinburgh. September 4–7 2006. V. 1. P. 469–476.
7. Elinas P., Sim R., Little J.J. SigmaSLAM: stereo vision SLAM using the rao-blackwellised particle filter and a novel mixture proposal distribution // IEEE Inter. Conf. on Robotics and Automation (ICRA). Orlando, Florida. May 15–19 2006. P. 1564–1570.
8. Fischler M.A., Bolles R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Communications for Association for Computing Machinery (ACM). 1981. V. 24. № 6. P. 381–395.
9. Georgiev Y. E90 Project: Stereo Visual Odometry: PhD thesis. Swarthmore College, 2006.

10. *Grest D., Krüger V., Koch R.* Single view motion tracking by depth and silhouette information // *Image Analysis*. 2007. V. 4522. P. 719–729.
11. *Huhle B., Jenke P., Strasser W.* On-the-fly scene acquisition with a handy multi-sensor system // *Inter. J. of Intelligent Systems Technologies and Applications*. 2008. V. 5. P. 255–264.
12. *Horn B.K.P.* Closed-form solution of absolute orientation using unit quaternions // *JOSA*. 1987. V. 4. № 4. P. 629–642.
13. *Hartley R., Zisserman A.* Multiple View Geometry in Computer Vision. Cambridge University Press, 2004. 672 p.
14. *Konolige K., Agrawal M., Bolles R.C., Cowan C., Fischler M., Gerkey B.P.* Outdoor mapping and navigation using stereo vision // *Proc. Inter. Symp. on Experimental Robotics (ISER)*. Rio de Janeiro, Brazil. July 6–10 2006. P. 122–131.
15. *Klein G., Murray D.* Parallel tracking and mapping for small ar workspaces // *Inter. Symposium on Mixed and Augmented Reality*. Nara, Japan. November 14–17 2007. P. 225–234.
16. *Lepetit V., Fua P.* Monocular model-based 3D tracking of rigid objects // *Found. Trends. Comput. Graph. Vis.* 2005. V. 1. № 1. P. 1–89.
17. *Lucas B.D., Kanade T.* An iterative image registration technique with an application to stereo vision // *Proc. Imaging Understanding Workshop*. Monterey, CA. November 1998. P. 121–130.
18. *Lange R., Seitz P., Biber A., Schwarte R.* Time-of-flight range imaging with a custom solid state image Sensor // *Proc. SPIE*. 1999. V. 3823. P. 180–191.
19. *Nuechter A.* 3D Robotic Mapping. The Simultaneous Localization and Mapping Problem with Six Degrees of Freedom // *Springer Tracts in Advanced Robotics*. 2009. V. 52. 204 p.
20. *Nister D., Naroditsky O., Bergen J.* Visual odometry // *IEEE Conf. on Computer Vision and Pattern Recognition*. Washington, DC. June 27 – July 2 2004. P. 652–659.
21. *Nuechter A., Surmann H., Lingemann K., Hertzberg J., Thrun S.* 6D SLAM with an application in autonomous mine mapping // *IEEE Inter. Conf. on Robotics and Automation*. New Orleans, LA, USA. April 26 – May 1 2004. P. 1998–2003.
22. *Parvizi E., Jonathan Wu Q.M.* Real-time 3D head tracking based on time-of-flight depth sensor // *IEEE Inter. Conf. on Tools with Artificial Intelligence*. October 29–31 2007. V. 1. P. 517–521.
23. *Rosten E., Drummond T.* Machine learning for high-speed corner detection // *European Conf. on Computer Vision*. Graz, Austria. May 2006. V. 1. P. 430–443.
24. *Rusinkiewicz S., Levoy M.* Efficient variants of the ICP algorithm // *3<sup>rd</sup> Inter. Conf. on 3D Digital Imaging and Modeling*. Quebec City, Canada. May 2001. P. 145–152.
25. *Streckel B., Bartczak B., Koch R., Kolb A.* Supporting structure from motion with a 3D-range-Camera // *Image Analysis*. 2007. V. 4522. P. 233–242.
26. *Shi J., Tomasi C.* Good features to track // *IEEE Conf. on Computer Vision and Pattern Recognition*. June 21–23 1994. P. 593–600.
27. *Thrun S., Burgard W., Fox D.* Probabilistic Robotics (Intelligent Robotics and Autonomous Agents). The MIT Press, 2005. 667 p.
28. *Zhang Z.* Iterative point matching for registration of free-form curves and surfaces // *Inter. J. of Computer Vision*. 1994. V. 13. № 2. P. 119–152.
29. *Zhang Z.* A flexible new technique for camera calibration // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000. V. 22. № 11. P. 1330–1334.