Humanoid Robot Dense RGB-D SLAM for Embedded Devices

Poster · April 2019

DOI: 10.13140/RG.2.2.13008.76803

5 authors, including:

Stylianos Piperakis
Foundation for Research and Technology - Hellas

Emmanouil Hourdakis

Dimitrios Kanoulas
Istituto Italiano di Tecnologia

Panos E. Trahanias
Foundation for Research and Technology - Hellas

人型机器人进行地图重建过程中出现姿态抖动，的解决方案

Humanoid Robot Dense RGB-D SLAM for Embedded Devices*

Stylianos Piperakis[1], Nikolaos Tavoularis[1], Emmanouil Hourdakis[1], Dimitrios Kanoulas[2] and Panos Trahanias[1]

***Abstract*— Visual Simultaneous Localization and Mapping (visual SLAM) constitutes a challenging task when applied to humanoid robots. While walking, the robot's feet strike the ground and generate sudden accelerations due to rapid and sequential contact switching. These in turn give rise to visual motion blurriness that greatly compromises the performance of the system. In this work we present a dense visual SLAM approach that integrates information from IMU, robot kinematics and contact measurements, to overcome these issues.**

I. Introduction

Localization is an essential piece of information for autonomous mobile robots, that is usually computed directly using the kinematic information of the platform. For humanoids, however, the kinematics often produces inaccurate estimates since they do not account for the dynamic effects caused by slippage, discontinuous ground contacts and actuation errors. In this context, Visual SLAM can provide an off-the-shelf method to compute the state of a humanoid accurately, using a light and low-cost sensor that is easily mountable on the robot.

However, the motion of a camera mounted on a humanoid has distinctive differences to the motion models assumed in traditional visual SLAM systems. It has a much wider spread, compared to the one for wheeled robots, and follows the oscillating trajectory of the center of mass, as designated by the bipedal gait. This results in blurriness during the image acquisition process, and reduced performance of the image registration methods.

II. Related work

To solve the aforementioned problems, research focused into new ways of integrating visual SLAM algorithms to humanoid robots [1]. Kagami et al. [2] proposed a vision-based, full-body motion planning method that fuses vision with the localization modules of a humanoid robot. In [3], visual information and a simplified dynamic model of the HRP-2 robot are utilized to improve the robustness of pose tracking. In [4] the authors investigated the performance of a monocular visual SLAM system on the NAO humanoid for which, due to hardware limitations, a robust visual SLAM algorithm was intractable. To overcome those limitations, the authors in [5] mounted an RGB-D sensor to NAO in order to integrate depth information to the robot's footstep planning. In [6], to estimate NAO's pose, a sparse visual SLAM system is employed along with an Extended Kalman Filter, based on kinematic and inertial information. More recently, dense visual SLAM methods are gaining interest, due to resilience to environments with poor features. The authors in [7] developed a dense RGB-D SLAM system for the HRP-4 humanoid, utilizing a reconstruction method suited for dynamic human environments. Finally, in [8] a dense visual SLAM method, with a semi-dense mapping, was embedded on NASA's Valkyrie humanoid.

*This work has been partially supported by the Greek National grand MIS 5030440 vipGPU.

[1]Stylianos Piperakis, Nikolaos Tavoularis, Emmanouil Hourdakis, and Panos Trahanias are with the Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH), Heraklion, Greece. {spiperakis,tavu, ehourdak,trahania}@ics.forth.gr

[2]Dimitrios Kanoulas is with the Humanoids and Human-Centered Mechatronics Department, Istituto Italiano di Tecnologia (IIT), Via Morego 30, 16163, Genova, Italy. Dimitrios.Kanoulas@iit.it.

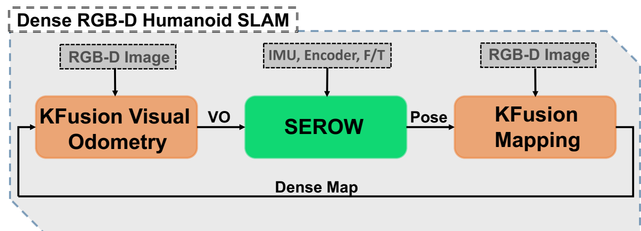

Fig. 1. Dense RGB-D Humanoid SLAM architecture. The framework utilizes RGB-D images along with IMU, encoder, and F/T measurements in an consistent loop.

However, the performance of dense visual SLAM, when computational power and sensor quality are limited, still remains an open question. In the current paper we present a robust RGB-D dense SLAM framework, based on Kinect-Fusion [9], [10], that effectively considers the humanoid's kinematics, feet contact status, and IMU measurements to: a) accurately estimate the robot's pose during locomotion, b) construct a dense map of the environment. Our implementation is offered as an open-source ROS C++ and CUDA package at www.github.com/tavu/kfusion_ros and achieves real-time execution on an Nvidia Jetson TX2 mounted on a NAO robot.

III. Method and Results

To efficiently consider the robot's kinematics and contact effects we employ the State Estimation Robot Walking (SEROW) framework [11], [12], [13]. The latter fuses IMU, joint encoder, and Force/Torque (F/T) measurements to accurately estimate the following state vector $\boldsymbol{x}_t$:

$$\boldsymbol{x}_t = \begin{bmatrix} {}^b\boldsymbol{v}_b & {}^w\boldsymbol{R}_b & {}^w\boldsymbol{p}_b & \boldsymbol{b}_{\boldsymbol{\omega}} & \boldsymbol{b}_{\boldsymbol{a}} \end{bmatrix}^\top$$

速度 旋转 位置 IMU的bias

where ${}^w\boldsymbol{p}_b$, ${}^w\boldsymbol{R}_b$ denote the base position and rotation with respect to a world frame w, ${}^b\boldsymbol{v}_b$ is the linear velocity, and $\boldsymbol{b}_{\boldsymbol{\omega}}$, $\boldsymbol{b}_{\boldsymbol{\alpha}}$ are IMU biases, in the base frame b.

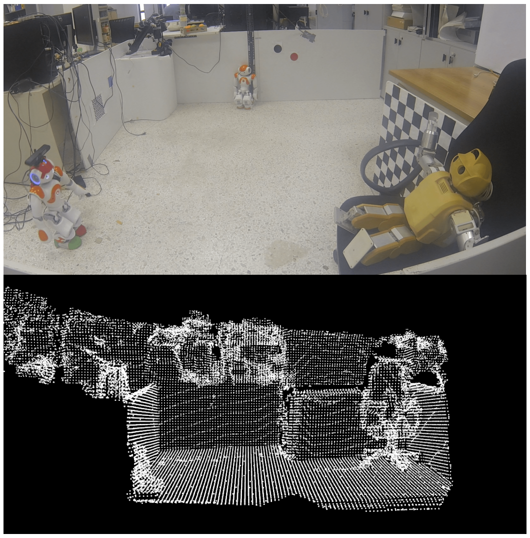

Fig. 2. Top: NAO navigating in an office environment. Bottom: Reconstructed 3D map.

The kinematic information derived from SEROW are used to improve the performance of the localization and mapping processes in visual SLAM. To accomplish this we modify KinectFusion [9] and separate two functionalities: a) Estimate the robot's pose using VO and b) Generate a dense map of the environment and rectify pose estimates using this map. In such it is straightforward to add SEROW in the loop, as illustrated in Fig. 1. SEROW incorporates the VO to estimate a more accurate robot pose that is then propagated to the mapping module.

We evaluate our system by executing it on a NAO humanoid, while navigating in an office environment (Fig. 2 top). Figure 3 demonstrates the corresponding 3D base position (Fig. 3, top) and orientation (Fig. 3, bottom) as estimated by KinectFusion, kinematics and KinectFusion with SEROW respectively. As it is shown, the estimates provided by KinectFusion alone exhibit large errors in the vertical axis, and are subject to drift. The latter is further illustrated in Table I that presents the final pose drift for the three implementations. Evidently, the fusion of KinectFusion with SEROW yields accurate estimates for all quantities, and improves the quality of the obtained map, as depicted in Fig. 2 bottom. The conducted experiment is available in HD quality at https://youtu.be/mLdNwHl9cgo. Our implementation achieved 29Hz on the Jetson TX2 module, given that RGB-D images are available at 30fps and volumetric rendering is disabled.

IV. Conclusion

In this paper, we presented a method to fuse a dense visual SLAM algorithm with the kinematic information from a humanoid robot. Our framework exhibits increased performance when compared to a solely RGB-D SLAM and can achieve real-time execution on an embedded GPU device.

TABLE I

Pose Drift

	${}^{w}p_x(m)$	${}^{w}p_y$	${}^{w}p_z$	roll (deg)	pitch	yaw
KF	0.025	0.047	0.287	2.80	2.35	2.86
Kin	0.261	0.331	0.023	6.5e-4	0.65	7.01
KF+S	0.011	0.038	0.031	6.5e-4	0.65	3.69

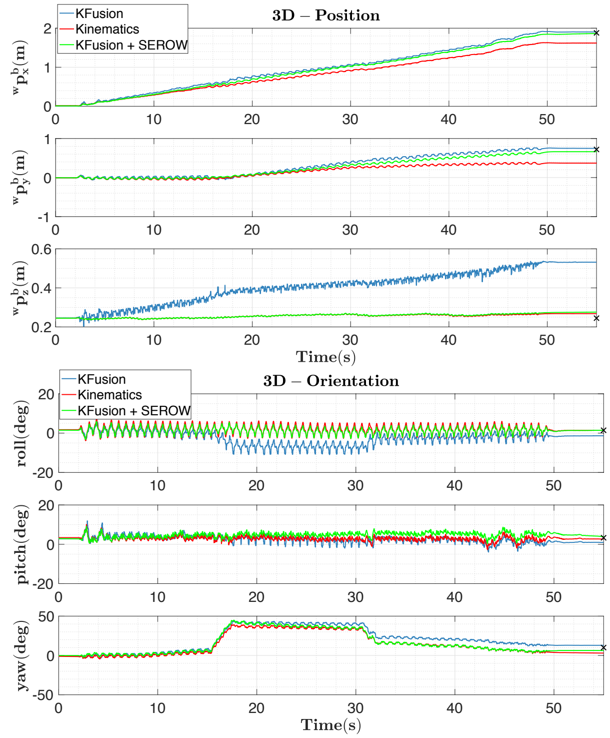

Fig. 3. Pose estimates generated by the three algorithms. Top: 3D body position. Bottom: 3D body orientation. Blue lines indicate the estimates from VO, red lines from the kinematics, green lines from KinectFusion and SEROW. Black crosses are the measured final states.

References

[1] W. Gouda, W. Gomaa, and T. Ogawa, "Vision based SLAM for Humanoid Robots: A Survey," in *IEEE JEC-ECC*, 2013, pp. 170–175.

[2] S. Kagami *et al.*, "Vision-based 2.5D Terrain Modeling for Humanoid Locomotion," in *IEEE ICRA*, 2003, pp. 2141–2146.

[3] O. Stasse *et al.*, "Integrating Walking and Vision to Increase Humanoid Autonomy," *IJHR*, vol. 5, no. 02, pp. 287–310, 2008.

[4] E. Wirbel, B. Steux, S. Bonnabel, and A. de La Fortelle, "Humanoid Robot Navigation: from a Visual SLAM to a Visual Compass."

[5] D. Maier, C. Lutz, and M. Bennewitz, "Integrated Perception, Mapping, and Footstep Planning for Humanoid Navigation among 3D Obstacles," in *IEEE/RSJ IROS*, 2013, pp. 2658–2664.

[6] G. Oriolo, A. Paolillo, L. Rosa, and M. Vendittelli, "Vision-based Odometric Localization for Humanoids using a Kinematic EKF," in *12th IEEE-RAS ICHR*, 2012, pp. 153–158.

[7] T. Zhang, E. Uchiyama, and Y. Nakamura, "Dense RGB-D SLAM for Humanoid Robots in the Dynamic Humans Environment," in *IEEE-RAS 18th ICHR*, 2018, pp. 270–276.

[8] R. Scona, S. Nobili, Y. R. Petillot, and M. Fallon, "Direct Visual SLAM Fusing Proprioception for a Humanoid Robot," in *IEEE/RSJ IROS*, 2017, pp. 1419–1426.

[9] S. Izadi *et al.*, "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera," in *24th Annual ACM Symposium on UIST*, 2011, pp. 559–568.

[10] D. Kanoulas, N. G. Tsagarakis, and M. Vona, "rxKinFu: Moving Volume KinectFusion for 3D Perception and Robotics," in *IEEE-RAS 18th ICHR*, 2018, pp. 1002–1009.

[11] S. Piperakis and P. Trahanias, "Non-linear ZMP based State Estimation for Humanoid Robot Locomotion," in *IEEE-RAS Intl. Conf. on Humanoid Robots*, 2016, pp. 202–209.

[12] S. Piperakis, M. Koskinopoulou, and P. Trahanias, "Nonlinear State Estimation for Humanoid Robot Walking," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3347–3354, Oct 2018.

[13] S. Piperakis, S. Timotheatos, and P. Trahanias, "Unsupervised Gait Phase Estimation for Humanoid Robot Walking," in *IEEE Intl. Conf. on Robotics and Automation*, 2019.