In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), LNCS 8681, pp. 797-804, Springer Heidelberg. Hamburg, DE, September 15-19, 2014

RatSLAM on Humanoids - A Bio-Inspired SLAM Model Adapted to a Humanoid Robot

Stefan Müller, Cornelius Weber, and Stefan Wermter

University of Hamburg, Department of Computer Science,
Vogt-Koelln-Strasse 30, D-22527 Hamburg - Germany
{3smuelle,weber,wermter}@informatik.uni-hamburg.de
http://www.informatik.uni-hamburg.de/WTM/

Abstract. Mapping, localization and navigation are major topics and challenges for mobile robotics. To perform tasks and to interact efficiently in the environment, a robot needs knowledge about its surroundings. Many robots today are capable of performing simultaneous mapping and localization to generate own world representations. Most assume an array of highly sophisticated artificial sensors to track landmarks placed in the environment. Recently, there has been significant interest in research approaches inspired by nature and RatSLAM is one of them. It has been introduced and tested on wheeled robots with good results. To examine how RatSLAM behaves on humanoid robots, we adapt this model for the first time to this platform by adjusting the given constraints. Furthermore, we introduce a multiple hypotheses mapping technique which improves mapping robustness in open spaces with features visible from several distant locations.

Keywords: SLAM, visual SLAM, RatSLAM, Humanoid robot, Mapping, Localization

1 Introduction

For successful and efficient interaction with the environment, world knowledge is needed. The challenge to gain this information can be addressed by the tasks of mapping, localization and navigation. Basic world interaction approaches rely on a-priori generated maps of static environments and perform localization with (noisy) odometric data and pre-defined landmarks. The main disadvantage of these approaches is the inability to deal with changes in dynamic environments, since geometry is not reliable to determine landmarks with single sensors [11] or arrays of different sensors [4].

During the last decades, many approaches tried to overcome this problem by creating an internal world representation for the robots themselves. The most successful ones form an entire group of methods that can perform Simultaneous Localization and Mapping (SLAM). Many combine multiple sensors, resulting in a multiplicity of sensor data that needs to be processed. Recently, SLAM has been used more frequently with humanoid robots to form probabilistic robot pose

In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), LNCS 8681, pp. 797-804, Springer Heidelberg. Hamburg, DE, September 15-19, 2014

estimates in complex 3D environments (SE3) supported by laser rangefinders and idiothetic on-board sensors [6] or to create accurate grid maps in SE2 [13].

An alternative to these approaches with high demand on processing power can be found in nature itself: Most animals do not have precise sensors like laser rangefinders to measure distances with an accuracy of millimeters for large ranges. Nevertheless, they perform successful mapping, localization and navigation tasks. Plenty of approaches try to adapt biological mechanisms and sensor usage to computational models. Most research in this area is focused on understanding the function of the brain and to develop exact biological models, tested only regarding their biological plausibility [1, 5]. Only few models, like from Arleo et al. [2] or Weiller et al. [14], have been tested regarding practical mapping, localization and performance on real robots. However, they have only been shown to work under many constraints in relatively small worlds.

In 2004, Milford et al. developed RatSLAM [10], a biological SLAM approach based on mapping and localization mechanisms in the rodent's hippocampus. In contrast to other models, biological validity and correctness was not as important as to create a reliable SLAM system with low computational complexity, usable on robots with low computing power. RatSLAM had been developed with and for wheeled robots. However, the physical characteristics and constraints of this type of robot differ from those of humanoids. Hence, applied to a humanoid, we encountered several problems related to robot instability, sensors and movement characteristics during evaluation. These issues led to false localizations and false positive loop closures in the topological map. To address these challenges, we developed an improved extended approach, called *Multi-hypotheses Experience Maps* (multi-EMs). This approach tracks multiple spatial robot position hypotheses at the same time and weights their plausibility, achieving more robust, less fault-prone mapping.

1.1 RatSLAM - A Bio-Inspired SLAM Solution

RatSLAM is a vision-based, biologically inspired model, able to achieve competitive SLAM results in real-world environments with a camera sensor and optionally sensors that gather odometric data [8, 9, 12]. Commonly used sensors like laser, ultrasonic or depth sensors are not used. It is a rough computational model of the part of the rodent's hippocampus that maintains its believed location in the world. RatSLAM uses techniques of landmark sensing in combination with odometric information to form a *Competitive Attractor Network* (CAN). This CAN forms a topological representation of adjacent world locations, which mostly includes Cartesian properties. RatSLAM consists of several processing units which are introduced here in short in the order they perform sensor data processing (fig. 1).

Local View (LV): The *local view* (LV) is a collection of simple neural units that store image templates. Templates are generated from a down-sampled 8bit-greyscale part (*Region of Interest*, ROI) of the current raw camera image. Templates are used to determine the robot location in space via scanline intensity

In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), LNCS 8681, pp. 797-804, Springer Heidelberg. Hamburg, DE, September 15-19, 2014

profile matching. Whenever a new camera image is received, the ROI is processed to a template and compared to all previously stored templates. If the new template sufficiently matches a stored one, the robot is deemed at a familiar location and the new template is not added to the network. Otherwise, if no sufficient match was found, the actual template is added to the LV cells for recognition.

Visual Odometry (VO): Odometric data is necessary to maintain an approximate robot location hypothesis if no visual cue is available. Usually, it is measured with sensors like rotary encoders [8, 9, 12]. Apart from that, RatSLAM is able to use visual methods to determine translational and rotational robot movement. In addition to the ROI for the LV, other ROIs are defined in the image: Forward movement and orientation changes are determined based on the rate of filtered average absolute intensity difference between consecutive scanline intensity profiles.

Pose Cell (PC) Network: The *Pose Cell network* is the core of RatSLAM and it forms three-dimensional localization and orientation hypotheses (x', y', θ') for the robot's pose within the real environment (x, y, θ). This network consists of a three-dimensional CAN of inter-connected neural units (PCs) with wrap-around connections. Each PC represents a location and orientation in the environment and is linked to LV cells by Hebbian learning links. The robot's current pose belief is represented by an activity level of the PCs. Cell activity can change due to injected energy whenever a familiar visual template is recognized. Multiple LV templates can match the same template and lead to a conformity level that is larger than a given threshold for multiple LV cells. All these LV cells inject energy into the PC network via the weighted Hebbian links. This can result in multiple activity packets being active at the same time. The total amount of energy, however, is kept constant by internal CAN attractor dynamics. The packet with the highest amount of energy is the strongest believed robot position. Another factor influencing the activity of PCs is path integration: The activity is shifted relative to odometry to nearby PCs while the robot moves in order to maintain consistency between real world and the internal map. Over time, the energy of packets can increase or decrease, new packets can appear, existing disappear or they can unite. For this reason, the robot's position cannot be determined for sure, and pose estimation is threated probabilistically.

Experience Map (EM): Experiments on the PC map showed that, especially in large environments, the PC representation is not topologically correct and only partially Cartesian [9, 12]. Reasons for this are path integration, from increasing odometric drift and particularly ever increasing numbers of re-localization ("loop closure") based on LV cells, linked to multiple world locations ("(hash) collision") and vice versa ("discontinuity"). Therefore, Milford et al. extended RatSLAM with a topological Cartesian world representation called *Experience Map* (EM) [8]. This map represents each world location by a unique experience $e_i = (P_i, L_i, \mathbf{p}_i)$ at an independent spatial position $\mathbf{p}_i = (x, y)$. The experience is linked to an individual PC $P_i(x', y', \theta')$ and exactly one LV cell L_i and gets

In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), LNCS 8681, pp. 797-804, Springer Heidelberg. Hamburg, DE, September 15-19, 2014

activated whenever the linked PC and the corresponding LV cell are active. Consecutive experiences are connected relative to each other by transitions which span a traversable graph that, in combination with information about the relative pose of involved experiences, movement behavior and movement duration for inter-experience traveling, can be used for path planning and navigation. A map correction algorithm inside the EM maintains Cartesian consistency at all time by relative location correction of experiences to each other to eliminate inconsistencies, which becomes obvious in loop closure events.

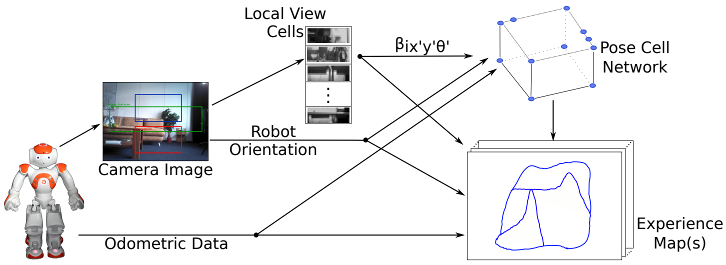

Fig. 1. The complete system consists of several processing units which in the end form a topological world representation inside the experience map.

2 Approach: RatSLAM on Humanoids

RatSLAM is available in an open source implementation called OpenRatSLAM [3] which has been used as basis for this adaption to humanoid robots. In the following, we will describe how RatSLAM is adapted to this type of robot. Thereafter, an extension is introduced to enhance the EM's overall accuracy.

2.1 Adapting RatSLAM to a Humanoid

RatSLAM was developed for wheeled robots. Hence, constraints have been made regarding the possible actions the robot can perform. Humanoid robots, however, have a different physical structure and come along with other constraints. These have to be integrated into the RatSLAM model while old constraints can be dismissed. To enable functionalities not implemented for wheeled robots, like forward, backward and sideways walking or turning on a spot, sensor data processing has been adapted: For our approach, the algorithm uses camera images for template generation and rotation detection only. Translational movements are obtained by a translational motion controller.

Humanoids in comparison to wheeled robots, move quite slowly while most of the image movement comes from the shaking of the robot during walking. Hence, exploration with humanoids takes much more time. To enhance exploration and

In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), LNCS 8681, pp. 797-804, Springer Heidelberg. Hamburg, DE, September 15-19, 2014

the overall map quality, we integrated an autonomous exploration approach, based on touch-and-turn techniques to create a spatial and comprehensive map. Over time, this exploration behavior spans a graph of trajectories over the complete environment that can be used for later navigation tasks. From time to time, the robot pauses and looks around on the spot, to create *anchor-points* which represent locations in the environment with independent view templates for plenty of different orientations. These locations extend the robot's narrow field of view (FOV) and improve re-localization capabilities immensely. We did not extend RatSLAM to account for the fact that humanoids can climb stairs and thereby may have access to 3D Euclidean space (SE3).

2.2 Multi-Hypotheses Experience Maps

One major issue while using RatSLAM is the affection to false positive loop closures during exploration ("*perceptual aliasing*"). False positive matches and snaps introduce inconsistency and irremediable failures to the map and have to be avoided. Generally, false positives are more serious than false negatives, as false negatives only reduce the overall recall rate but in the end have negligible impact on the total map precision [7]. This difficulty is caused by two reasons:

1. Robot instability during movement: A relatively high center of mass in combination with a comparatively high-mounted camera on the robot's forehead results in swaying movements whenever the robot moves. VO based on image differences does not work under these conditions;
2. Image quality: Due to the low image resolution, the LV algorithm is unable to distinguish locations with almost identical orientations but different distances to the same environmental feature and assumes an identical position.

Although multiple pose hypotheses in the PC network and threshold adaption for image classification introduce stability, false positive loop closures appear frequently. Further increased thresholds would lead to many missing (false negative) loop closures.

In our approach, we introduce multiple EMs with multiple robot pose hypotheses at once to increase the overall accuracy and to repair hastily loop closures (fig. 1). Independent loop closures for each EM create different hypotheses for traveled paths and the current robot pose. All EMs get ranked in comparison to an artificial map of arithmetic means $\overline{em}$ for all experiences e_i in all EMs k (with $k \geq 2$) linked to the same PC pc_i:

$$\bar{x_i} = \frac{1}{n}\sum_{k=1}^{n} x_{i_k} \qquad \bar{y_i} = \frac{1}{n}\sum_{k=1}^{n} y_{i_k} \tag{1}$$

The rating is done for the last n experiences. Regarding figure 2, the distance d_i of two experiences $\overline{em}_i$ and e_i is calculated by

$$d_i(\bar{e_i}, e_i) = \sqrt{(\bar{x_i} - e_{i,x})^2 + (\bar{y_i} - e_{i,y})^2} \quad = \sqrt{(\Delta x + \Delta y)^2}. \tag{2}$$

In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), LNCS 8681, pp. 797-804, Springer Heidelberg. Hamburg, DE, September 15-19, 2014

Experiences existing for some time have already adjusted their pose by the path integration algorithm. Their position is more reliable than recently created experiences. Hence, to strengthen already modified positions, each distance d_i is weighted dependent on the time the experience exists.
The overall aberration $\boldsymbol{\Delta}(\overline{em}, em_k)$ of EMs em_k and $\overline{em}$ considering the last n experiences with $n \leq |em_k|$ is computed by

$$\boldsymbol{\Delta}(\overline{em}, em_k) = \frac{1}{n} \sum_{i=|em_k|-n}^{|em_k|} (|em_k| - i)\sqrt{(\Delta x_i + \Delta y_i)^2} \tag{3}$$

Periodically, based on $\boldsymbol{\Delta}(\overline{em}, em_k)$, the EM with the highest accumulated x- and y-distance values is rated worst. If one map was rated worst for four times, it is replaced by a copy of the best ranked EM. The more frequent this replacement is performed, the more the system does rely on its odometry in place of re-localizations.

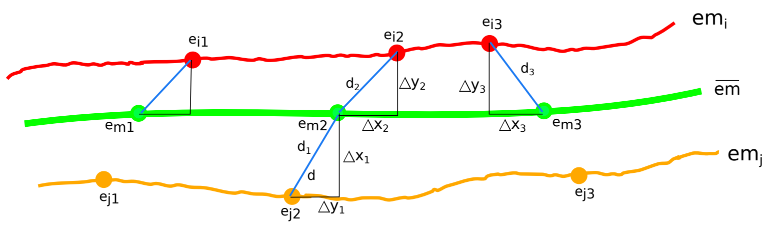

Fig. 2. The distance d_i for an experience to a calculated experience e_{m_i} is based on the x, y-offset. Adjusted weights for d_i refer to the age of the experience.

3 Experiments

The modified RatSLAM system has been tested on a NAO[1] robot. This 58cm tall humanoid is equipped with cameras with narrow Field-of-Views (FOVs, HOR: 60.9°, VER: 47.6°) located in its head. 14 joint motors offer 25 DOFs for flexible movements. RatSLAM and the NAO were linked through a *Robot Operating System*[2] (ROS) wrapper[3] to make the robot's API accessible with ROS.

SLAM was performed in a domestic environment with daylight from a window. RatSLAM was confronted with ambiguous situations that, normally, lead to false re-localizations and loop closures. Anchor-points (yellow spots in fig. 3), created during exploration, represent locations with visual templates for different orientations. All paths were planned in a way that locations near these points are traversed more than once and map ranking was triggered every two seconds.

[1] http://www.aldebaran-robotics.com/
[2] http://www.ros.org/
[3] http://wiki.ros.org/Robots/Nao

In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), LNCS 8681, pp. 797-804, Springer Heidelberg. Hamburg, DE, September 15-19, 2014

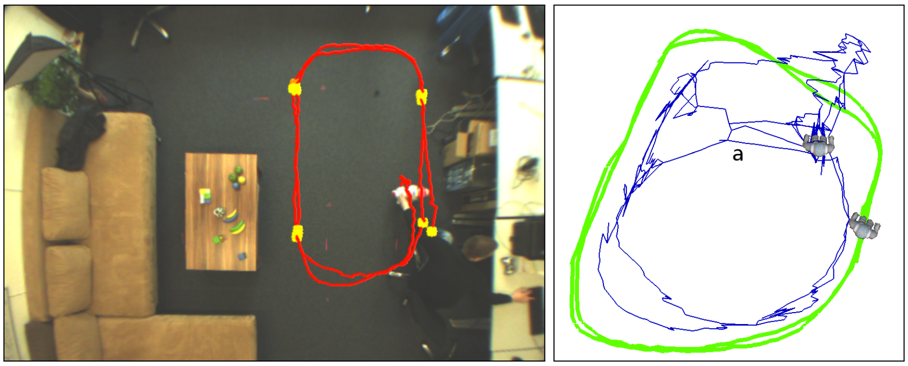

Fig. 3. The robot walked the red path of $2m \times 4m$ with yellow anchor-points twice. The internal map for one EM is drawn in thin blue (a defect resulting from false snapping is marked by "a"), the mean-map based on 4 EMs in bold lime-green color.

4 Results and Discussion

We evaluated the mapping and localization abilities of RatSLAM on the NAO robot, especially for the crucial task of loop closure. The integration of anchor-points enhances the recognition rate of familiar locations. Although it cannot achieve the accuracy of a bi-directional path exploration, it is a benefit for mapping tasks on humanoids with slow movement speeds and limited horizontal FOV. The focus of this work is on the enhancement of the system to Multi-Experience-Maps to reduce false positive loop closures and to strengthen the map's accuracy. To generate comparable results, we tested the system's original implementation as well as the modified version with multiple EMs on identical data sets. Fig. 3 displays the walked trajectory (red) as automatically recorded by a ceiling camera with anchor-points (yellow), the internal map based on the original RatSLAM approach (blue) and based on a map of means $\overline{em}$ (green).

Results show that this extension reduces technical drawbacks of the humanoid architecture. As can be seen in figure 3, it interpolates the jagged trajectory by correcting the location of experiences. Together with the replacement of the worst rated EM, $\overline{em}$ prevents the whole system from hasty false loop closures in ambivalent situations (like at "a") and, over time, leads to a more accurate world representation without false connections between distant experiences.

5 Conclusion

So far, RatSLAM has only been used in combination with wheeled and aerial robots. Our approach adapted RatSLAM to humanoids. The different physical architecture of this type of robot led to different robot characteristics and constraints and, hence, did not allow using the model without adjustments and new constraints. This led to problems that differ a lot from the ones on previous implementations. Therefore, the used NAO robot responded with swaying

In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014), LNCS 8681, pp. 797-804, Springer Heidelberg. Hamburg, DE, September 15-19, 2014

movements and blurred camera images with lost details and washed-out features. Independent of the internal parameters of RatSLAM this led to multiple false positive loop closures. Our new approach creates multiple pose estimates in several EMs and deals with false loop closures. This improves the overall topological map structure and therefore increases the accuracy of the map representation which can be used for navigation and many further tasks that include navigation. Hence, the extension of RatSLAM with multi-EMs is now a model usable on humanoid robots as well.

References

1. A. Arleo, J. A. del, and D. Floreano. Efficient Learning of Variable-Resolution Cognitive Maps for Autonomous Indoor Navigation. In *IEEE Trans. Robot. Automat*, pages 990–1000, 1999.
2. A. Arleo and W. Gerstner. Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biological cybernetics*, 83(3):287–299, Sept. 2000.
3. D. Ball, S. Heath, J. Wiles, G. Wyeth, P. Corke, and M. Milford. OpenRatSLAM: an open source brain-based SLAM system. 34(3):149–176, 2013.
4. J. A. Castellanos, J. Neira, and J. D. Tardos. Multisensor fusion for simultaneous localization and map building. *Rob. and Autom., IEEE Trans. on*, 17(6):908–914, Dec. 2001.
5. M. Franz and H. A. Mallot. Biomimetic Robot Navigation. In *Robotics and autonomous Systems*, pages 133–153, 2000.
6. A. Hornung, K. M. Wurm, and M. Bennewitz. Humanoid robot localization in complex indoor environments. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ Int. Conf. on*, pages 1690–1695. IEEE, Oct. 2010.
7. M. Magnusson, H. Andreasson, A. Nüchter, and A. J. Lilienthal. Automatic appearance-based loop detection from three-dimensional laser data using the normal distributions transform. *Journal of Field Robotics*, 26(11-12):892–914, 2009.
8. M. Milford, D. Prasser, and G. Wyeth. Experience mapping: producing spatially continuous environment representations using RatSLAM. In *Conf. on Rob. and Autom.*, 2005.
9. M. Milford, G. Wyeth, and D. Prasser. Efficient Goal Directed Navigation using RatSLAM. In *Rob. and Autom.. ICRA. Proc. of the 2005 IEEE Int. Conf. on*, pages 1097–1102. IEEE, Apr. 2005.
10. M. J. Milford, G. F. Wyeth, and D. Prasser. RatSLAM: a hippocampal model for simultaneous localization and mapping. In *Proc. of the IEEE Int. Conf. on Rob. and Autom. (ICRA)*, volume 1, pages 403–408. IEEE, Apr. 2004.
11. S. Nuske, J. Roberts, and G. Wyeth. Robust Outdoor Visual Localization Using a Three-Dimensional-Edge Map. 2009.
12. D. Prasser, G. Wyeth, and M. Milford. Experiments in outdoor operation of ratslam. In *Proc. of Australasian Conf. on Rob. and Autom.*, 2004.
13. C. Stachniss, M. Bennewitz, G. Grisetti, S. Behnke, and W. Burgard. How to learn accurate grid maps with a humanoid. In *Rob. and Autom., 2008. ICRA. IEEE Int. Conf. on*, pages 3194–3199. IEEE, May 2008.
14. D. Weiller, L. Läer, A. K. Engel, and P. König. Unsupervised learning of reflexive and action-based affordances to model adaptive navigational behavior. *Frontiers in Neurorobotics*, 4, 2010.