2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)
Beijing, China, November 6-9, 2018

Effects of Biped Humanoid Robot Walking Gaits on Sparse Visual Odometry Algorithms

Yukitoshi Minami Shiguematsu[1], Martim Brandao[2], Kenji Hashimoto[3], and Atsuo Takanishi[4]

Abstract— **Motivated by experiments showing that humans regulate their walking speed in order to improve localization performance, in this paper we explore the effects of walking gait on biped humanoid localization. We focus on step length as a proxy for speed and because of its ready applicability to current footstep planners, and we compare the performance of three different sparse visual odometry (VO) algorithms as a function of step length: a direct, a semi-direct and an indirect algorithm. The direct algorithm's performance decreased the longer the step lengths, which along with the analysis of inertial and force/torque data, point to a decrease in performance due to an increase of mechanical vibrations. The indirect algorithm's performance decreased in an opposite way, i.e., showing more errors with shorter step lengths, which we show to be due to the effects of drift over time. The semi-direct algorithm showed a performance in-between the previous two. These observations show that footstep planning could be used to improve the performance of VO algorithms in the future.**

Index Terms - **Localization, ego-motion, visual odometry, humanoid robot, WABIAN-2R**

I. INTRODUCTION

If humanoid robots are to become more useful in our daily life, their capabilities should be improved to be able to autonomously realize various tasks effectively and in a robust manner. The ability to self-localize in the environment is an important step in that direction. One common way for the robot to self-localize is through odometry algorithms, i.e., the estimation of the robot's change in position through the use of motions sensors, such as cameras, inertial measurement units (IMU), motor encoders, etc. These sensors can be used independently, as is the case of visual odometry (VO) algorithms, or their information can be combined to get better estimates, using algorithms as the Kalman or particle filters.

One approach to improve localization performance is to change the path a robot takes to a goal or the goals themselves in a way that optimizes said performance.This is called active localization, which refers to the act of partially or fully controlling the motions of the robot to minimize the uncertainty and increase the efficiency and robustness of the estimation of its current pose [1], [2]. Humanoid robots could potentially change inter-limb coordination or gait parameters while keeping the same base trajectory, to affect the camera motion and improve the robot's localization performance.

One example of the effect of walking styles on self-localization systems can be found in humans, as we regulate our walking speed in order to improve our localization with our eyes closed [3]. However, current humanoid robot walking controllers and localization systems are built in ways fundamentally different from that of biological systems, and are not built purposely to achieve similar localization performance behavior (i.e. similar relationship between walking speed and localization accuracy). Moreover, previous work with hexapod robots has found inconclusive and irregular variation of SLAM performance with gait parameters [4].

With the above in mind, the contribution of this paper is to answer the following questions regarding localization systems for biped humanoid robots:

- Does performance of such systems depend consistently and non-trivially with humanoid gait?
- What effects do different walking styles have on the performance of such systems?

The approach in this paper is data-driven, i.e., we do not try to predict localization performance from simplified mechanical, control, sensor, or environment models. Instead, we directly measure localization performance of the whole system, by using ground-truth data from motion capture from experiments while varying the robot's walking gait parameters. For this paper, the parameter we focus on is step length, as it is a useful and readily applicable representation for humanoid robot locomotion planning, as in footstep planning, and because several relevant humanoid robot locomotion performance metrics as energy and slippage have been shown to depend on step length [5]. We describe this data-based approach and the data analysis in Section III. In section IV we present the relationships found between localization performance for three different VO algorithms and the mentioned gait parameters in our robotic platform. We also we discuss the possible explanations for the observed relationships.

II. RELATED WORK

Self-localization for humanoid robots has been widely researched. In the case of VO algorithms, Stasse et al. [6] proposed a real-time monocular Visual Simultaneous Localization and Mapping (VSLAM) algorithm taking into account robot kinematics from the walking pattern generator.

*This study was conducted as part of the Research Institute for Science and Engineering, Waseda University, and as part of the humanoid project at the Humanoid Robotics Institute, Waseda University. It was also supported in part by the Program for Leading Graduate Schools, the Graduate Program for Embodiment Informatics of the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT, Japan), by SolidWorks Japan K.K and Cybernet Systems Co.,Ltd. M. Brandao is funded by UK Research and Innovation and EPSRC, ORCA research hub (EP/R026173/1).

[1] Graduate School of Science and Engineering, Waseda University, Tokyo, Japan. contact@takanishi.mech.waseda.ac.jp

[2] Oxford Robotics Institute, University of Oxford, UK.

[3] School of Science and Technology, Meiji University; and is a researcher at the Humanoid Robotics Institute (HRI), Waseda University.

[4] Department of Modern Mechanical Engineering and Humanoid Robotics Institute (HRI), Waseda University, Tokyo, Japan.

In [7], an IMU based state estimation for a stereo based 3D SLAM is proposed, using measurements from the stereo VO and robot kinematics as updates for the Extended Kalman Filter. In [8], the authors propose an unscented Kalman filter to estimate the ankle and hip states of a biped robot, to then use a support vector regression learning controller for bipedal walking. Xinjilefu et al. [9] propose a decoupled estimation, first using a joint dynamics estimator, and then a base link position estimator, instead of including all that information in a single filter, in order to reduce the computational cost but sacrificing some accuracy. In [10], a bipedal robot state estimator is proposed, based on another originally designed for a quadruped robot [11]. These estimators make the filter update based on feet measurements.

Regarding active localization using vision sensors, Davidson et al. [12] were the first to take the effects of actions into account for localization, using a stereo system attached to a mobile robot and trying to minimize the motion drift along a predefined trajectory. Also, one of the common approaches is the "Next Best View" approach, which as the name states, seeks a single additional sensor placement to reduce the localization error of the system [13]. There have been works proposing different criteria to estimate the influence of the robot motions on SLAM, for example focusing on Kalman filter based approaches [14], or on the effect of the camera motion on the stability of visual localization for aerial robots [15].

Also for aerial robots, methods to select paths with minimum pose uncertainty while considering the robot's dynamics have been proposed [16], as well as methods to plan paths with richer visual features [17], and more recently a method that computes the localization uncertainty optimally incorporating photometric and geometric information [2].

For legged robots, [4] assessed the localization accuracy of a hexapod robot in different types of terrain changing the robot's gait accordingly, using an RGB-D sensor. For biped humanoid robots, active visual localization has been researched from different perspectives, as active localization to improve the interactions of the robot with its environment for object manipulation [18], an active vision system to estimate the location of objects while walking [19], or a task-oriented active vision system for a vision-guided bipedal walking [20]. However, none of the above assessed the effect of the walking motion itself on the performance of the robot's localization, nor used this information to plan or modify the walking gait of the robot to obtain a better localization estimate.

From the biological point of view, humans mainly use visual, gravitational/inertial and proprioceptive cues for ego-motion estimation. Moreover, humans change the weight they put on each sensory input depending on the situation [21], [22]. Also, we know humans plan their walking gait ahead in many situations, as to keep stability in difficult situations like slippery terrains [23], but we also change gait parameters when there are problems with the sensory inputs, such as decreasing walking speed or leaning the trunk backwards when visual disturbances arise [24]. Moreover, different sensory modalities perform better depending on our motions. The visual system performs better at lower frequencies than the vestibular system, but both are integrated in an optimal manner [25]. There is also evidence pointing out that modifying the walking speed has effects on our path integration abilities, making us underestimate distances when walking at slower speeds [3], as well as walking cadence affecting the performance of path integration, achieving the best performance at about 2 Hz [26].

III. METHODOLOGY

As explained in Section I, in this paper we focus on the effects of step length on localization performance. We generated one walking pattern for each step length with a fixed total walking distance of 1.5 m on a straight line. Step lengths were 0.1, 0.125, 0.15, 0.175, 0.2, 0.225 and 0.25 m. Step lengths shorter than 0.1 m resulted in unstable gaits, while step lengths longer than 0.25 m were not tested since they were close to the mechanical limits of the robot. The step width was maintained constant at 0.08 m. Five runs were performed for each step length with the robot having the knees bent, i.e., maintaining a fixed height for the center of mass (CoM), and therefore for the camera. The reference walking cadence was fixed to 0.96 s/step, 0.06 seconds for double support phase and 0.9 seconds for single support phase. All patterns were executed on the robot by joint position control without any state estimation (i.e. assuming the reference trajectory of the base was executed perfectly). The motion capture and robot's joints, force, IMU and image data were stored and later analyzed [1]. The experiments were carried out inside a texture-rich laboratory environment.

A. System Overview

For the experiments in this paper we used the biped humanoid robot WABIAN-2R [27] (Fig. 1), a 33 Degrees of Freedom (DoF) bipedal humanoid robot. For the visual input, we used a Matrix Vision mvBlueCOUGAR-X, a global shutter monocular camera, together with a low distortion wide angle lens of focal length 1.28 mm, a Field of View (FOV) of 125 deg and a distortion of 3%. The stream of images was set to 117 Hz, and the camera was mounted on the head of the robot (Fig. 2). For the ground truth measurements, a motion capture system OptiTrack V120:Trio at 120 fps was used, placing the photo-reflective markers on the camera to obtain the actual trajectory.

The different reference frames and transformations used for the experiments can be seen on Fig. 3. We use two main reference frames, W and C_t, the world frame and the camera system frame at time t, respectively. Also, following the notation used in [28], we define ${}^{(est)}T_{A_{t_i} \to B_{t_j}}$ as the transformation of frame B at time t_j relative to frame A at time t_i, calculated with the estimator est. The motion capture system tracks the camera system in the world frame, ${}^{(gt)}T_{W \to C_t}$, whereas the VO system tracks the motion of the camera system relative to its initial frame, ${}^{(vo)}T_{C_{init} \to C_t}$.

[1]The datasets can be found at:
https://github.com/yuchango/wabian_experiments

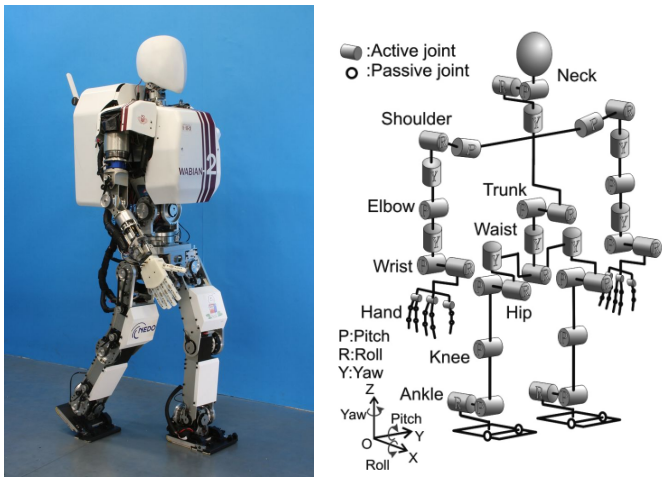

Fig. 1. Robotic platform WABIAN-2R (left) and DoF configuration (right).

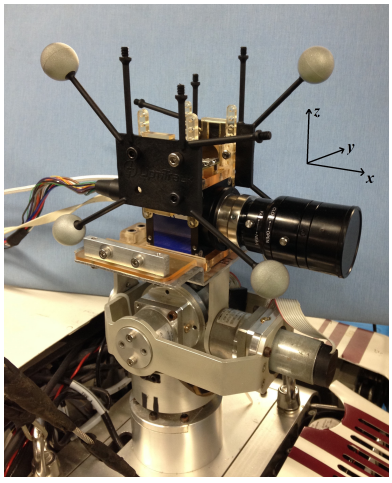

Fig. 2. Close-up of the head system used for localization and ground-truth (head, camera, reflective markers)

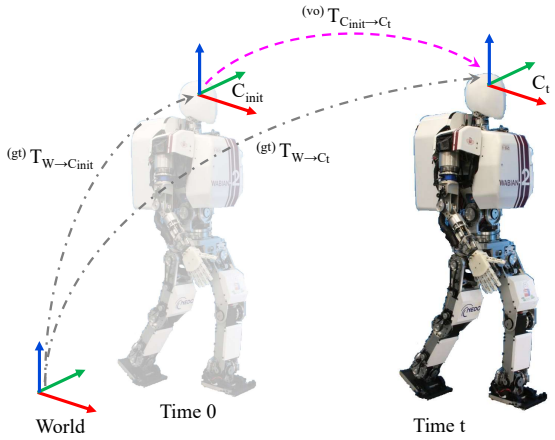

Fig. 3. Used coordinate frames.

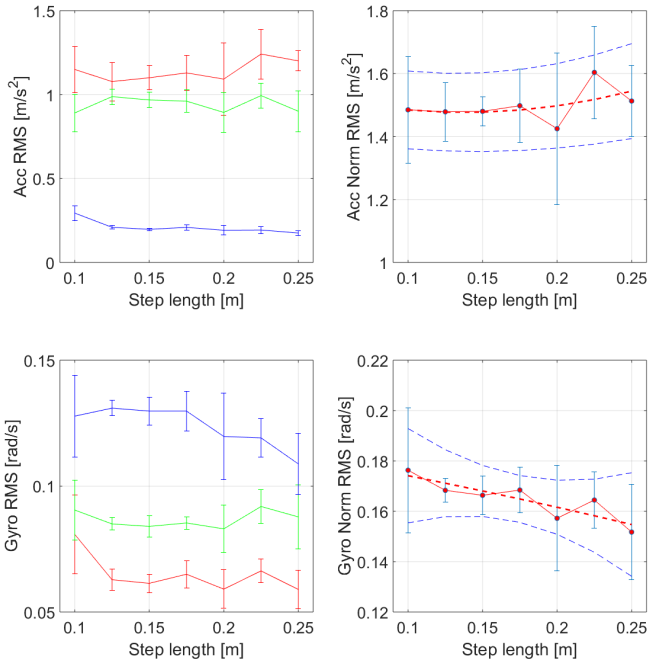

Fig. 4. RMS of the data from the accelerometer and gyroscope of the IMU mounted on the camera for x (red), y (green) and z (blue) (right), and their norm, with a fitted quadratic curve (left).

For the visual localization, we tested three state of the art monocular visual odometry algorithms: SVO 2.0 [29], ORB-SLAM2 [30] and DSO [31], which we treated as black boxes, and will be briefly explained in the next section. We fed the image stream and the intrinsic parameters of the camera, and extracted the estimated position and orientation of the camera.

We also logged acceleration and angular velocity data at 200 Hz from one IMU mounted on the camera itself, as well as force and torque data from sensors placed on both feet, also at 200 Hz. This data was processed and analyzed to look for possible differences between different walking speeds (Figs. 4, 5).

B. Visual Odometry Algorithms

Visual odometry algorithms are normally classified based on two properties. Depending on the information they use from the images, the algorithms can be *Direct* or *Indirect*. *Direct* methods use pixel intensity information, making them fast, but prone to errors caused by changes in lighting conditions. *Indirect* methods pre-process the raw sensor data, extracting features such as corners, edges, or more sophisticated feature descriptors, which makes them more robust to lighting changes, but computationally heavier because of the feature calculation process. On the other hand, depending on the amount of pixels used from the input images, the algorithms are classified as *Dense* or *Sparse*. *Sparse* methods use and reconstruct only a selected set of independent points, whereas *Dense* methods attempt to use and reconstruct all pixels in the 2D image domain.

For this paper, we decided to test *Sparse* methods, as we are focusing on localization and do not need to reconstruct a map from the visual input. Therefore, we selected a *Direct* and *Sparse* method (DSO) [31], a *Semi-Direct* and *Sparse*

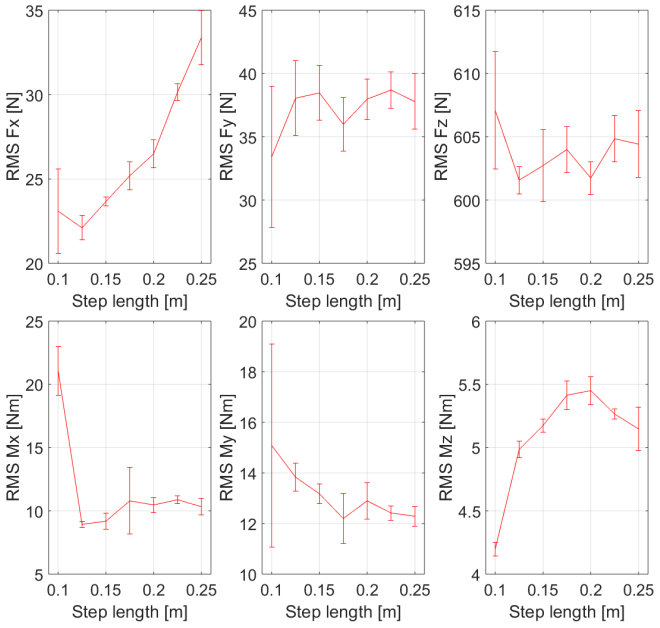

Fig. 5. RMS of the data from the F/T sensors on the robot's feet.

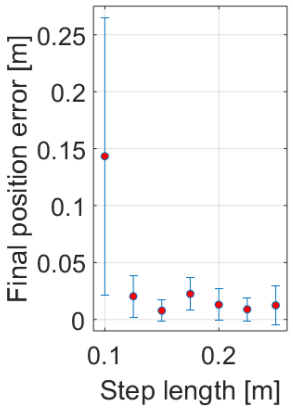

Fig. 6. Trajectory tracking error, obtained with the motion capture system, w.r.t. the reference motion trajectory.

method (SVO 2.0) [29] and an *Indirect* and *Sparse* method (ORB-SLAM2) [30].

C. Scale Extraction

To solve the scale ambiguity problem of monocular localization algorithms, for each VO algorithm and for each experiment we calculated the scale from the estimated traveled distance of the camera after the first step and from the reference step length. Assuming a flat floor and no slipping on the first step:

$$\lambda_{est} = \frac{^{(ref)}d_{first\ step}}{^{(vo)}d_{first\ step}} \tag{1}$$

where λ_{est} is the estimated scaling factor, and $d_{first\ step}$ is the Euclidean distance between the initial position of the camera system and its position after the first step. We also chose this method as it is one of the hypothesized ways in which humans try to calculate traveled distances while walking, using substratal idiothetic cues, i.e., based on information about movement with respect to the ground or to inertial space [32].

For comparison, we also used the scale calculated using the actual traveled distance obtained from the ground truth:

$$\lambda_{real} = \frac{^{(gt)}d_{first\ step}}{^{(vo)}d_{first\ step}} \tag{2}$$

to examine how close was with respect to the estimated scale with respect to the real one.

IV. DATA ANALYSIS

For the analysis of the localization performance of the different VO algorithms, we focused on the absolute trajectory error (ATE), and the relative pose error (RPE) [33]. Both are calculated after aligning the trajectories using the method of Horn [34], which finds the rigid-body transformation corresponding to the least-squares solution that maps the estimated trajectory onto the ground truth trajectory in closed form.

The ATE is used to asses the global consistency of the estimated trajectory, by comparing the absolute distances between the estimated and the ground truth trajectories, after both trajectories have been aligned (Fig. 7).

$$ATE_t = {}^{(gt)}\mathrm{T}^{-1}_{W\to C_t}\,{}^{(vo)}\mathrm{T}_{W\to C_t} \tag{3}$$

We then evaluated the root mean squared error over all time stamps of the translational components:

$$RMSE(ATE_t) = \left(\frac{1}{n}\sum_{i=1}^{n}\|ATE_i\|^2\right)^{\frac{1}{2}} \tag{4}$$

On the other hand, the RPE is used to asses the drift between the estimated and ground truth trajectories. We set the time interval Δ to 10 [ms], assuming that in this time interval the motion is linear (Fig. 8).

$$RPE_t = {}^{(gt)}\mathrm{T}^{-1}_{C_t\to C_{t+\Delta}}\,{}^{(vo)}\mathrm{T}_{C_t\to C_{t+\Delta}} \tag{5}$$

Similar to the ATE, we evaluate the root mean squared error over all time stamps, with $m = n - \Delta$:

$$RMSE(RPE_t) = \left(\frac{1}{m}\sum_{i=1}^{m}\|RPE_i\|^2\right)^{\frac{1}{2}} \tag{6}$$

A. Discussion

The results from the walking experiments showed an interesting relationship between the visual localization accuracy and the robot's step lengths used to cover the 1.5 m trajectory. The estimation from SVO 2.0 resulted to be the one with the least error, followed by ORB-SLAM2 and DSO. Interestingly, DSO seems to be affected the most by accelerations, possibly vibrations on the camera caused by the walking motions. As the step length increases, both the acceleration and the ATE of DSO increase (Fig. 4, 7). However, in Fig. 7 (top) the error for a step length of 10 cm

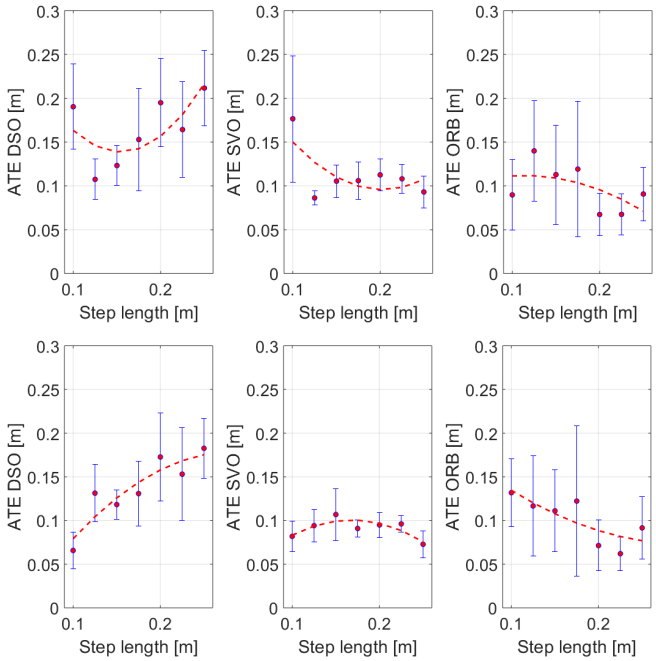

Fig. 7. ATE versus step length, for DSO, SVO 2.0 and ORB-SLAM2 for λ_{est} (top) and λ_{real} (bottom). Red dots with blue vertical error bars denote the average and standard deviations for each step length, while the red dashed lines are the fitted quadratic curves for the averages. Fitted quadratic curves were calculated using the $polyfit$ function of MATLAB ®.

seemed like an outlier, which is also true for SVO 2.0. This coincides with the fact that with the same step length, forces in the vertical axis, as well as torques around x and y are the biggest (Fig. 5). This happens normally when there are early contacts of the feet with the ground or slippage, which makes the motions unstable. Also, looking at the trajectory execution error (i.e. the difference between the reference trajectory to reach the target 1.5 m away and the actual final position measured by the motion capture system), the error at a step length of 10 [cm] is the biggest (Fig. 6). Therefore, using λ_{est} for a step length of 10 cm would not be adequate, since the actual step length can be far from the actual step length. Changing the scale to λ_{real} highlighted the trend of the relationship between the performance of each VO algorithm and step length (Fig. 7, bottom). From this figure, we clearly see how DSO's error increases with step length, which besides coinciding with the accelerometer data, also coincides with an increase of contact forces on the feet in the x axis (Fig. 5).

In the case of ORB-SLAM2, the ATE decreases as the step length increases (Fig. 7). This could point that this algorithm, without optimizations or loop closures, is the most affected by drift. With smaller step lengths, the time to travel the reference distance is longer, which would then increase the effects of drift, whereas with longer step lengths the time is less, and so would be the drift, causing less estimation errors in the end. Finally, the trend of SVO 2.0 could be explained by the fact that, as it is a semi-direct algorithm, both effects from vibrations and drift are combining together.

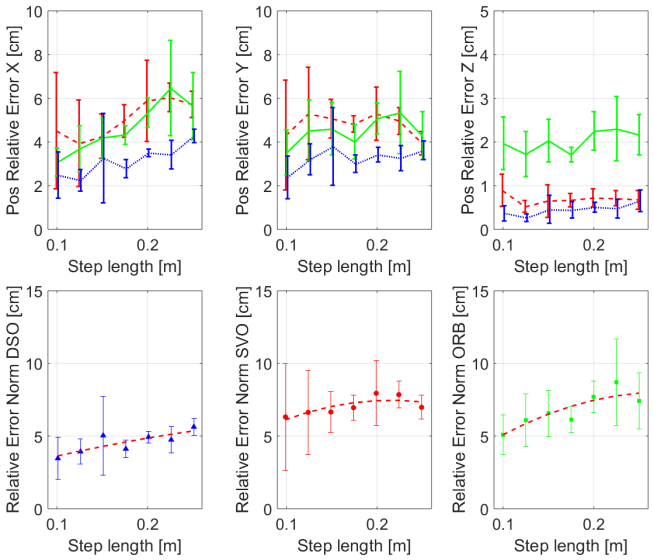

Fig. 8. Relative error of the different VO algorithms w.r.t. the ground truth. SVO 2.0 (red), ORB-SLAM" (green), DSO (blue).

V. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

We performed a set of experiments with a biped humanoid robot walking with different step length values, in order to find out whether this parameter would affect the performance of VO algorithms. As we are focusing on localization and not mapping algorithms, we tested a Direct (DSO), a Semi-direct (SVO2.0) and an Indirect (ORB-SLAM2) VO algorithm. Increasing the step length of the walking gaits showed an increase on the acceleration measurements, most likely because faster walking introduced more vibrations on the robot, which affected the performance of DSO. Also, we observed worse localization performances for SVO 2.0 and ORB-SLAM2 the shorter the step lengths, i.e., the slower the walking, and as it took more time to get to the goal, the effect of drift on the localization estimates was also increased. From the above, we observe that to minimize the effect of drift, we need to walk faster, but this in turn produces more vibrations, which negatively affect the localization performance. Therefore, this paper shows that footstep planning could be used in an active localization system to improve the performance of VO algorithms.

B. Future Works

As we observed a correlation between walking step length and visual localization performance, we are planning to include these performance curves as cost functions within a footstep planner such as to minimize localization error.

Regarding the calculation of the scale for monocular VO algorithms, in this paper we showed that obtaining it from an assumed step length leads to low performance. We are planning to explore ways of extracting the scale before starting the motion and/or during the motion itself, as is the case for some insects that present peering behaviors, or birds that use head-bobbing [35]. Experiments in environments

with more or less texture should also be carried out in order to investigate the influence of visual texture on the visual localization performance.

Also, we are interested in exploring how the localization performance is influenced by other gait parameters such as stepping time, i.e. the duration of single and double support phases. We are planning to explore as well the influence of gait in other (non-visual or fusion) localization algorithms.

REFERENCES

[1] D. Fox, W. Burgard, and S. Thrun, "Active markov localization for mobile robots," *Robotics and Autonomous Systems*, vol. 25, no. 3-4, pp. 195–207, 1998.

[2] G. Costante, C. Forster, J. Delmerico, P. Valigi, and D. Scaramuzza, "Perception-aware path planning," *arXiv preprint arXiv:1605.04151*, 2016.

[3] J. Bredin, Y. Kerlirzin, and I. Isral, "Path integration: is there a difference between athletes and non-athletes?" *Experimental Brain Research*, vol. 167, no. 4, pp. 670–674, Dec. 2005.

[4] J. Faigl, "On localization and mapping with RGB-D sensor and hexapod walking robot in rough terrains," in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 002 273–002 278.

[5] M. Brandao, K. Hashimoto, J. Santos-Victor, and A. Takanishi, "Footstep planning for slippery and slanted terrain using human-inspired models," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 868–879, Aug 2016. [Online]. Available: http://www.martimbrandao.com/papers/Brandao2016-tro.pdf

[6] O. Stasse, A. J. Davison, R. Sellaouti, and K. Yokoi, "Real-time 3d slam for humanoid robot considering pattern generator information," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 348–355.

[7] S. Ahn, S. Yoon, S. Hyung, N. Kwak, and K. S. Roh, "On-board odometry estimation for 3d vision-based SLAM of humanoid robot," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 4006–4012.

[8] L. Wang, Z. Liu, C. L. P. Chen, Y. Zhang, S. Lee, and X. Chen, "A UKF-Based Predictable SVR Learning Controller for Biped Walking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 6, pp. 1440–1450, Nov. 2013.

[9] X. Xinjilefu, S. Feng, W. Huang, and C. G. Atkeson, "Decoupled state estimation for humanoids using full-body dynamics," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 195–201.

[10] N. Rotella, M. Bloesch, L. Righetti, and S. Schaal, "State estimation for a humanoid robot," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 952–958.

[11] M. Bloesch, M. Hutter, M. A. Hoepflinger, S. Leutenegger, C. Gehring, C. D. Remy, and R. Siegwart, "State estimation for legged robots-consistent fusion of leg kinematics and IMU," *Robotics*, p. 17, 2013.

[12] A. J. Davison and D. W. Murray, "Mobile robot localisation using active vision," in *Proceedings of the 5th European Conference on Computer Vision - Volume II*, ser. ECCV '98. London, UK: Springer-Verlag, 1998, pp. 809–825.

[13] E. Dunn and J.-M. Frahm, "Next Best View Planning for Active Model Improvement." in *BMVC*, 2009, pp. 1–11.

[14] R. Sim and N. Roy, "Global a-optimal robot exploration in slam," in *2005 IEEE International Conference on Robotics and Automation*, 2005, pp. 661–666.

[15] C. Mostegel, A. Wendel, and H. Bischof, "Active monocular localization: Towards autonomous monocular exploration for multirotor mavs," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3848–3855.

[16] M. W. Achtelik, S. Lynen, S. Weiss, M. Chli, and R. Siegwart, "Motion- and Uncertainty-aware Path Planning for Micro Aerial Vehicles: Motion and Uncertainty Aware Path Planning," *Journal of Field Robotics*, vol. 31, no. 4, pp. 676–698, July 2014.

[17] S. A. Sadat, K. Chutskoff, D. Jungic, J. Wawerla, and R. Vaughan, "Feature-rich path planning for robust navigation of MAVs with mono-SLAM," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3870–3875.

[18] D. Gonzalez-Aguirre, M. Vollert, T. Asfour, and R. Dillmann, "Robust real-time 6d active visual localization for humanoid robots," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2785–2791.

[19] O. Lorch, J. F. Seara, K. H. Strobl, U. D. Hanebeck, and G. Schmidt, "Perception errors in vision guided walking: analysis, modeling, and filtering," in *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, vol. 2. IEEE, 2002, pp. 2048–2053.

[20] J. Seara and G. Schmidt, "Intelligent gaze control for vision-guided humanoid walking: methodological aspects," *Robotics and Autonomous Systems*, vol. 48, no. 4, pp. 231 – 248, 2004.

[21] R. J. Peterka, "Dynamic Regulation of Sensorimotor Integration in Human Postural Control," *Journal of Neurophysiology*, vol. 91, no. 1, pp. 410–423, Sept. 2003.

[22] H. van der Kooij, R. Jacobs, B. Koopman, and F. van der Helm, "An adaptive model of sensory integration in a dynamic environment applied to human stance control," *Biological cybernetics*, vol. 84, no. 2, pp. 103–115, 2001.

[23] R. Cham and M. S. Redfern, "Changes in gait when anticipating slippery floors," *Gait & Posture*, vol. 15, no. 2, pp. 159 – 171, 2002.

[24] A. Hallemans, E. Ortibus, F. Meire, and P. Aerts, "Low vision affects dynamic stability of gait," *Gait & Posture*, vol. 32, no. 4, pp. 547–551, Oct. 2010.

[25] F. Karmali, K. Lim, and D. M. Merfeld, "Visual and vestibular perceptual thresholds each demonstrate better precision at specific frequencies and also exhibit optimal integration," *Journal of Neurophysiology*, vol. 111, no. 12, pp. 2393–2403, June 2014.

[26] H. S. Cohen and H. Sangi-Haghpeykar, "Walking speed and vestibular disorders in a path integration task," *Gait & Posture*, vol. 33, no. 2, pp. 211–213, Feb. 2011.

[27] Y. Ogura, K. Shimomura, H. Kondo, A. Morishima, T. Okubo, S. Momoki, H. ok Lim, and A. Takanishi, "Human-like walking with knee stretched, heel-contact and toe-off motion by a humanoid robot," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, Oct 2006, pp. 3976–3981.

[28] R. Scona, S. Nobili, Y. R. Petillot, and M. Fallon, "Direct visual slam fusing proprioception for a humanoid robot," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2017, pp. 1419–1426.

[29] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "Svo: Semidirect visual odometry for monocular and multicamera systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, April 2017.

[30] M. J. M. Mur-Artal, Raúl and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[31] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, March 2018.

[32] M.-L. Mittelstaedt and H. Mittelstaedt, "Idiothetic navigation in humans: estimation of path length," *Experimental Brain Research*, vol. 139, no. 3, pp. 318–332, Aug. 2001.

[33] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2012, pp. 573–580.

[34] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *JOSA A*, vol. 4, no. 4, pp. 629–642, 1987.

[35] A. Bruckstein, R. J. Holt, I. Katsman, and E. Rivlin, "Head movements for depth perception: Praying mantis versus pigeon," *Autonomous Robots*, vol. 18, no. 1, pp. 21–42, 2005.