

Continuous Humanoid Locomotion over Uneven Terrain using Stereo Fusion

Pat Marion*, Maurice F. Fallon**, Robin Deits*, Thomas Whelan***, Matthew Antone*,
John McDonald**** and Russ Tedrake*

* Massachusetts Institute of Technology, Cambridge, USA

pmarion, rdeits, mantone, russt@mit.edu

** University of Edinburgh, UK

maurice.fallon@ed.ac.uk

*** Imperial College London, UK

t.whelan@imperial.ac.uk

**** Maynooth University, Ireland

johnmcd@cs.nuim.ie

1 Introduction

Humanoid robots have the potential to traverse the same complex terrain as humans: dealing with all the challenges of uneven and discontinuous surfaces, cracks and gaps while moving at typical human speeds. While locomotion research spans actuator development, dynamic planning and control, in this work we focus on terrain estimation and footstep planning — in particular while in continuous motion.

We demonstrate that a precisely accurate dense terrain map can be estimated using only imagery from a humanoid robot's *passive stereo camera*. This result then provides the necessary sensory input for an on-demand footstep planner which can in turn compute a set of footsteps which are optimal (within our assumptions) and kinematically feasible; yet does so sufficiently quickly that the robot can continuously locomote over uneven terrain.

Our experiments demonstrate the Boston Dynamics Atlas robot continuously walking over a terrain course similar to the DARPA Robotic Challenge (DRC) walking task. Our aim with this work is to demonstrate how modern dense visual mapping techniques can be used by dynamic robots while in motion.

2 Visual Stereo Fusion

Several previous research groups have explored the use of visual input in humanoid walking planning — typically by detecting distinct edges of clutter and obstacles rather than consciously detecting the actual terrain. In this work we estimate the underlying terrain using stereo fusion.

Two advances in visual mapping have enabled high quality dense 3D reconstructions at real-time. The first is high quality stereo cameras (paired with on-board FPGA block matching) which are robust enough to produce stereo point clouds at 15-30 Hz while in motion. The Atlas robot is equipped

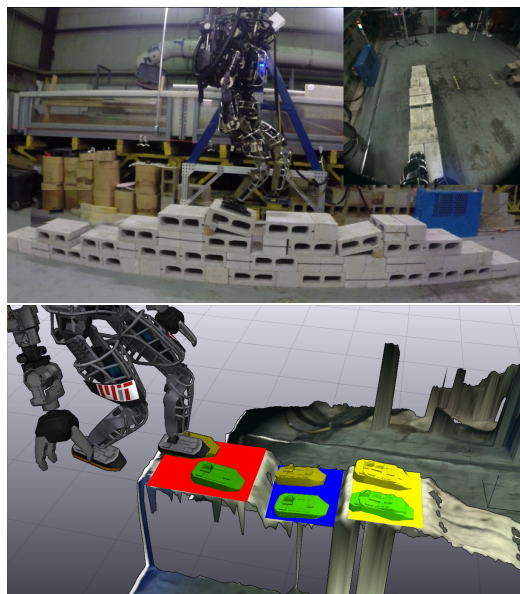


Figure 1: Visualization of the robot continuously locomoting over a complex terrain course. The upper figure shows the side view of the experiment and the view from the robot's stereo camera. The lower figure shows the constructed dense stereo map with the segmented walkable regions and the planned footsteps ready to be executed.

with an industrially hardened stereo camera manufactured by Carnegie Robotics.

The second advance is a new generation of dense visual fusion algorithms such as Kintinuous, [1], which use a GPU to integrate the noisy raw data into a fused de-noised map of the robot's vicinity in real-time.

As Kintinuous was originally developed for active RGB-D (e.g. Microsoft Kinect) data, significant pre-filtering is required to reject anomalous points estimated by the FPGA

block matching e.g. due to repeated texture. A video showing a comparison between raw stereo, fused stereo, and LIDAR, can be seen at:

<http://youtu.be/0ibv09D3JIw>

3 Terrain Segmentation

A terrain map is generated by the Kintinuous system each time the robot lifts its standing foot. We use a point cloud based segmentation algorithm to find locally planar convex regions of the terrain map for input to the footstep planner. (The segmentation algorithm is capable of processing sensor terrain data from either LIDAR or projected dense stereo maps.)

Our segmentation algorithm models the terrain course as planar, rectangular stepping blocks with arbitrary position and rotation. The first step is to remove all points that represent the floor (ground plane). Next, we apply point cloud surface normal estimation using a local search neighborhood of 5 cm around each point. The points are filtered to keep only those within 30 degrees of horizontal according to surface normals. Steeper regions are deemed infeasible for footsteps.

The filtered points are input to a Euclidean clustering routine that finds individual connected components of uniform planar segments. For each planar cluster we compute the minimum-area enclosing rectangle. A convex hull could also be used but in practice we used the bounding box so as to fill in missing areas that were filtered due to poor surface normal estimation (block edges and boundaries).

4 Footstep Planning

We use the mixed-integer convex optimization described by Deits and Tedrake in [2] to quickly produce optimal footstep plans given convex regions of safe terrain. Thus, when planning footsteps, the entire perception system can be abstracted away into a tool that produces regions of safe terrain. These regions, along with a navigation goal position are used as input to the footstep planning. We use the *carrot on a stick* approach to define a goal position a few meters ahead of the robot. The optimization chooses the number of footsteps to take and the poses of those footsteps in order to bring the robot close to its goal and thus makes progress across the terrain.

A key aspect of this work is the system integration which enabled the entire system to work in real-time. Our computation was shared between two identical off-board desktops each with an 3.30GHz Intel i7 CPU and an Nvidia GeForce GTX 680 GPU. The computation time of each step of the chain was as follows:

- Image acquisition: 105 msec (incl FPGA Matching)
- Speckle component removal: 40 msec
- Kintinuous stereo fusion: 110 msec (averaging 9.5Hz)

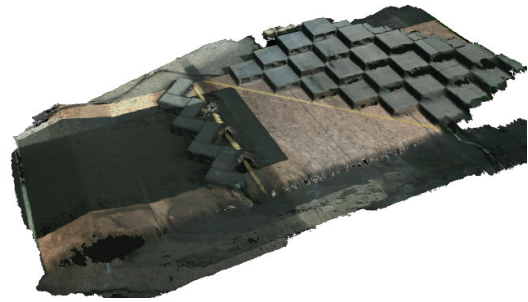


Figure 2: Reconstruction of the DRC terrain course (4-by-8m).

- Planar region segmentation: 615 msec (400-1100 msec)
- Footstep planning: 445 msec (300-600 msec)

Footsteps were executed by successively passing the steps to the manufacturer's controller for execution.

5 Experiments

In a series of progressively more complicated experiments we developed the capability until the robot could continuously cross the course illustrated in Fig. 1. The course was 5.5m in length and stretched the entire length of our test area. The robot autonomously walked over the course in 240 seconds for a total of 25 steps and 14 rows of blocks. A video showing the procedure and the experiment can be seen here:

http://youtu.be/_6WQxXH-bB4

We believe that this represents the first demonstration of general purpose visual reconstruction being used in combination with on-demand footstep planning to allow continuous locomotion over complex terrain courses.

Finally, we demonstrate in Figure 2 that the reconstruction works as well outdoors as inside using data collected by the robot outdoors during the walking task at the DRC Trials (Dec 2013).

An interesting direction for further work is direct stabilization of the sensor system by the robot's whole-body control algorithm. We would also like to implement a more general visual exploration strategy than was demonstrated here.

References

- [1] T. Whelan, M. Kaess, H. Johannsson, M. Fallon, J. Leonard, and J. McDonald, "Real-time large scale dense RGB-D SLAM with volumetric fusion," *Intl. J. of Robotics Research*, 2015, to appear.
- [2] R. Deits and R. Tedrake, "Footstep planning on uneven terrain with mixed-integer convex optimization," in *IEEE/RSJ Int. Conf. on Humanoid Robots*, Madrid, Spain, 2014.