Dynamic Locomotion Terrain Recognition for the MIT Cheetah 3 Robot 6.869 / 6.819: Advances in Computer Vision

Gerardo Bledt
Massachusetts Institute of Technology
Cambridge, MA
gbledt@mit.edu, 6.869

Milo Knowles
Massachusetts Institute of Technology
Cambridge, MA
mknowles@mit.edu, 6.819

Abstract

Our goal in this project is to create an inexpensive, lightweight and robust vision system that can estimate the terrain in front of the MIT Cheetah 3 robot in real time. Using a stereo rig consisting of two $20 webcams, we are able to obtain metrically accurate 3D reconstructions of planar obstacles from image pairs. In addition, by fitting a polytope to noisy terrain point clouds, our representation of the terrain is robust to outliers and has a very compact parametrization. This allows real time communication of the sensed enfironment to the robot, which allows it to modify its trajectory plan according to the vision input.

1. Introduction

Currently, the MIT Cheetah 3 robot uses primarily completely blind locomotion. What this means is that there is no vision or LiDAR system that allows it to preemptively sense its environment or plan for upcoming obstacles. Dynamic gaits such as trotting have been stabilized under ideal flat-terrain conditions, but most interesting applications of robotics involve some kind of unstructured, non-flat, non-smooth terrain situations. Recent research efforts have allowed for reactive tactile environment sensing where the robot is able to interpret and build its knowledge of the environment when it senses unexpected contact with a surface [1]. This has allowed for the same nominal dynamic gaits to stabilize in rough terrains by reacting to the estimated foot contacts.

However, since this environment sensing is purely reactionary, the robot can only detect objects and uneven ground after it has already determined contact with the surface. This covers a lot of cases where small objects and holes are littered over terrain with a slowly changing gradient. A current challenge is when the robot encounters a large, sharp terrain disturbances, such as a box or a set of stairs, it will not be able to smoothly use this reactionary sensing to navigate the terrain. If the box or stair is higher than the nominal foot swinging height as it is in Figure 1, it will simply not be able to step over it and will not detect a support surface contact when it does hit the side. Similarly, if the box or stair is higher than its possible foot swing height, it will not be able to continue forward and crash. By incorporating some visual terrain recognition, the robot will be able to modify its swing foot trajectory to clear objects that are within its kinematic workspace, as well as feed terrain information to the contact detection algorithm for more accurate probabilistic ground height models.

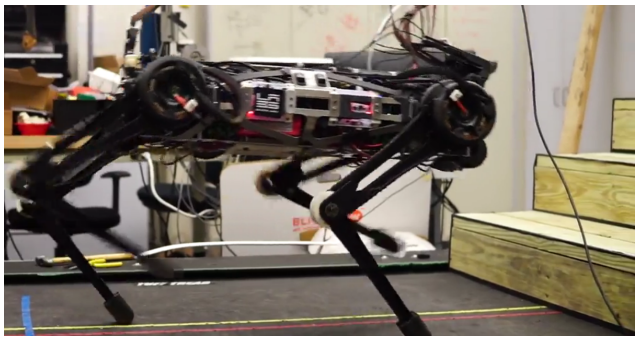

Figure 1. A robust controller is used for general robot locomotion in unstructured terrains, but a method of external environment sensing is needed for the robot to know that its nominal foot swing trajectory will not be able to clear the stair obstacle.

Paper on creating the disparity map [2]

2. Related Work

3. Approach

Our pipeline is summarized in Figure 2. We begin with a pair of RGB stereo images. From this image pair and calibration parameters obtained offline, we apply a stereo matching algorithm [See TODO] to get a depthmap of the scene. Using camera instrinsic parameters, we can project this depthmap into 3D space and obtain a point cloud of the environment [See TODO]. We segment locations in the left

RGB image that are likely candidates for planar surfaces, and sample corresponding points in the point cloud to obtain an estimate for the parameters of a plane. Finally, we construct a polytope of the object in front of the robot by joining the planar surfaces that we detect. The robot can then use this geometric representation of the terrain to plan its foot trajectory.

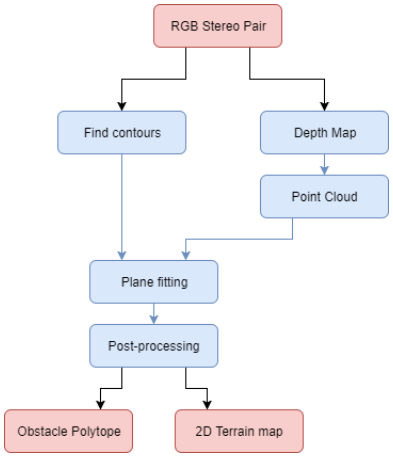

Figure 2. An overview of our obstacle reconstruction pipeline.

3.1. Stereo Matching

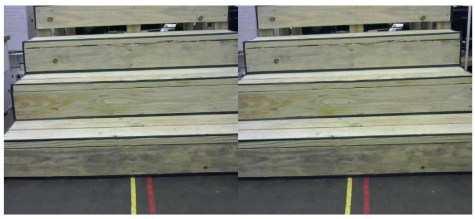

Figure 3. An example stereo image pair taken by our camera setup. The two cameras are parallel and have a baseline of approximately 8cm.

3.2. Point Cloud from Depthmap

3.3. Polytope Estimation of Objects

4. Results

Since the robot previously used only reactive sensing of the environment this meant that the only way for the robot to sense obstacles and terrain was to physically interact with them. A robust reactive controller has been achieved and allows the robot to traverse highly unstructured terrains while remaining stable. However, while this makes the robot robust, there are situations where the terrain is not traversable

Figure 4. An example stereo anaglyph.

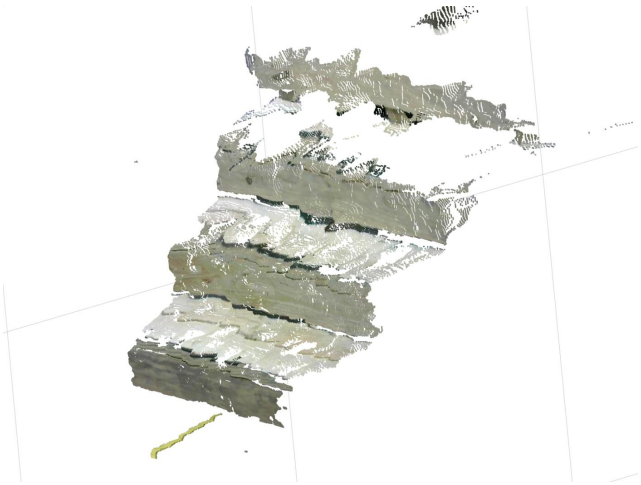

Figure 5. A pointcloud representation of the set of stairs in Figure 3.

easily without prior knowledge of the environment. Adding the vision system to the overall control system as an input to the path planning as seen in Figure 9 will allow the robot to know if an obstacle is present, where it is, and its dimensions. Then it can decide how to modify its locomotion plan accordingly. For purposes of this demo the robot will attempt to detect the stair height and distance and modify its swing leg trajectory to place it on the stairs as its nominal swing foot height is lower than an average step.

It is not enough to just find the dimensions of the object in front of the robot. This doesn't have any physical meaning to the robot until we find the relationship between the reconstructed object to the robot's coordinate frame. We take the points of interest of the object in the 3D camera frame, ${}^{\mathcal{C}}p_o$, and apply the following transformation

$$ {}^{\mathcal{I}}p_o = {}^{\mathcal{I}}H_{\mathcal{C}}\,{}^{\mathcal{C}}p_o \tag{1} $$

where ${}^{\mathcal{I}}H_{\mathcal{C}}$ is the homogeneous transformation matrix be-

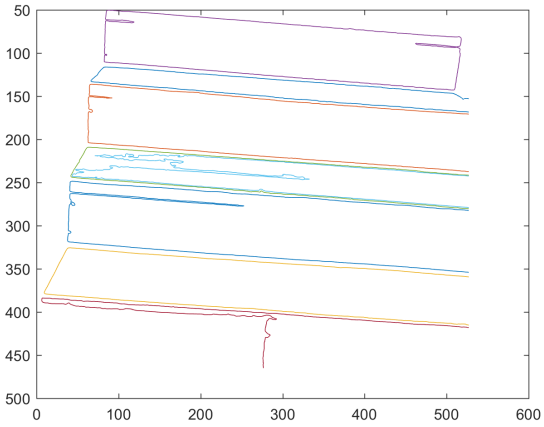

Figure 6. An example contour plot of the stairs in Figure 4. Here we have shown the eight largest contours in the image by area.

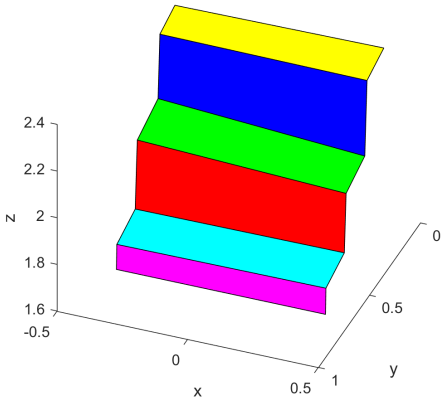

Figure 7. The output of our pipeline when applied to the stereo pair in Figure 3. The pink, red, and blue planes correspond to horizontal faces of the stairs. The light-blue, green, and yellow planes are vertical faces.

tween the camera frame and the Inertial frame. This allows the robot to know the distance and angle to the stairs relative to its body as seen in Figure 10. With this, the robot now has knowledge about its environment without the need to physically interact with it. With the robot holding an internal model of the environment with respect to its state, the path planner can adapt the nominal gait and foot placement trajectory to deal with the stairs at the detected height. The foot trajectory is modified when the robot's center of mass has traveled close enough to the detected stairs

$$({}^{\Psi}p_{CoM,x} - {}^{\Psi}p_{0,x}) < (d - d_{buffer}) \qquad (2)$$

where ${}^{\Psi}p_{CoM,x}$ is the yaw rotated CoM position ignoring pitch and roll effects, ${}^{\Psi}p_{0,x}$ is the yaw rotated position at the last vision input, d is the distance to the object at the

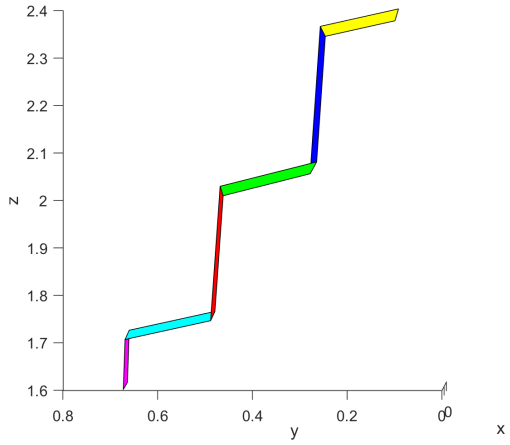

Figure 8. The polytope from Figure 7 viewed from the side. The vertical faces of this polytope (light-blue, green, yellow) are very close to the groundtruth dimension of 20cm. Similarly, the remaining horizontal faces are close to their groundtruth dimension of 31cm. Note that the pink plane is cut off by a bounding box around the polytope, so it does not appear to be the correct size.

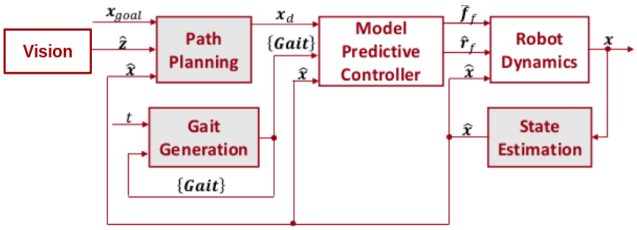

Figure 9. .

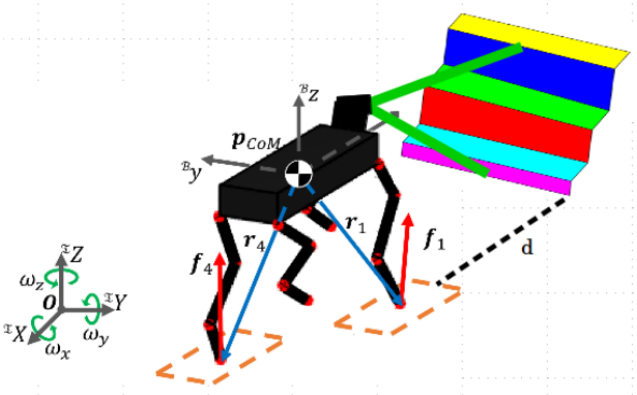

Figure 10. The robot must transform the coordinates of the stairs in the 3D reconstructed camera frame into the world frame in relation to the robot.

detection time, and d_{buffer} is a user defined buffer distance so that the robot begins lifting the legs slightly before the actual stair.

The vision system was successfully integrated with the robot software pipeline. Testing was first done in simulation where the webcams were used to detect the stairs and

the robot was walked forward in simulation. The nominal trotting swing height for the leg trajectories is about 15 cm and the stair height was 21 cm. Therefore, the path planner made the decision to raise the height of the trajectory to about 30 cm, which was approximately 10 cm over the detected object height. When it reached the location of the 3D reconstructed stairs, the robot modified its foot trajectory appropriately.

The pair of cameras was mounted on the physical robot and experiments were run with the real stair set on a treadmill. Figure 11 shows the robot successfully climbing the stairs on its first attempt. With the terrain-blind controller, the most common failure mode was the robot being unable to get its leg over the stair and tripping. With the vision input, this was no longer the case as it knew how high to lift its leg and where to place the feet to aquire an adequate foothold.

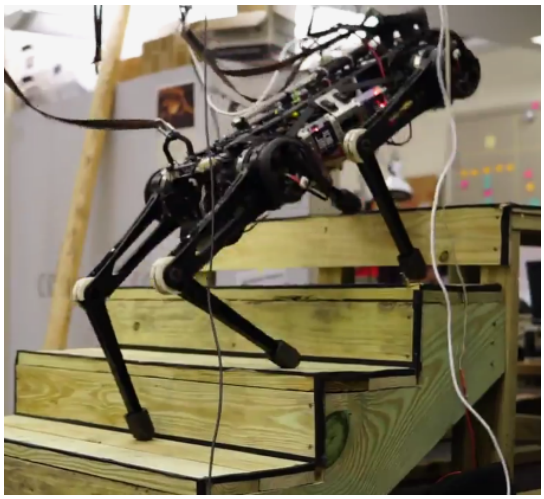

Figure 11. With the vision input, the robot is able to stably figure out how high to raise the feet in order to place them onto the stairs and successfully climb up.

5. Conclusion

The presented vision system yielded promising results for use as a general terrain sensing solution for the MIT Cheetah robot during locomotion. Successful experiments showed its ability to accurately detect the location and dimensions of the stairs. Terrain information being provided to the robot helps the robot's path planning algorithms make better decisions when facing obstacles. Stability is improved as it anticipates contacts more accurately, rather than blindly relying on its reactive interactions with the physical world.

However, several practical limitations were noted when implementing the system. First and most importantly, the results were extremely sensitive to the camera calibration. As soon as the cameras moved slightly from the position they were calibrated in, the algorithms would return poor reconstructions. Similarly, if there was a slight lag between the frames that were taken on each camera, it would essentially act as if the cameras had moved with respect to each other since the robot is constantly moving. The stereo algorithm also relies on having enough distinct features to detect the disparity between the images. While most real world situation will likely have enough features, it is another consideration to keep in mind. The work presented is a good first step towards robust robot vision, but more work would still be needed in the future. Overall, the system was successful in allowing the robot to climb stairs reliably under the experiment conditions.

Abstract (10Introduction (10Related work (10Approach (and technical correctness) (20Experimental results (and technical correctness) (20Conclusion (6.667References (3.333Overall clarity of the report (10Reproducibility: can the work be reproduced from the information given in the report? (10

References

[1] G. Bledt, P. Wensing, S. Ingersoll, and S. Kim. Contact model fusion for event-based locomotion in unstructured terrains. *(submitted) 2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[2] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 807–814 vol. 2, June 2005.