Visual system for a Modular 6-Legged robot

Kartik Patath, Hadi Salman, Howie Choset

***Abstract*— Search and rescue has always been a critical question to be answered in times of natural disasters. The destroyed areas are difficult and sometimes impossible for humans to access. Biologically inspired legged robots can be quite effective at search and rescue operations, especially in confined areas where humans cannot reach. These robots can reduce risk levels of human rescuers. This creates an opening to new research area in the field of search and rescue. In most of the scenarios the legged robots work using compliance, for motion in unknown environments. We propose a vision system for a 6-legged robot which enables it to better perceive the environment, helping the robot to navigate in an unknown environment.**

I. Introduction

Search and rescue takes many forms, each with its own unique risks and dangers to victim and responder. Urban search and rescue has become familiar to many of us as recent natural and man-made disasters have made the news all over the world. In some situation such as wilderness search and rescue there will often be only one victim with a large geographic search region. There are also numerous recent mine collapses in the United States and China [1]. These rescue operations are mostly performed by humans. There are significant dangers rescuers face during search and rescue tasks in a uniquely hazardous operating environment, such as mines for example. In case of destroyed buildings, there are many voids that are simply too small for people and dogs to enter, thus limiting the search to no more than a few feet from the perimeter. Robots on the other hand can bypass the danger and expedite the search for victims immediately after a collapse. This advanced technology can be useful to the rescue workers in ways such as: (1) reducing personal risk to workers and dogs by entering unstable structures, (2) allowing to penetrate ordinarily inaccessible voids, (3) increasing the efficiency by searching areas with multiple sensors, to provide a complete search in three dimensions, and (4) extending the reach of rescuers to go places that were otherwise inaccessible.

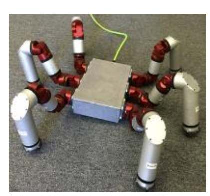
In this paper, we consider a platform for doing search and rescue known as the Snake Monster. It is basically a 6-Legged robot. The Snake Monster is reconfigurable robot with reliable field-tested hardware. By using parallel position and force control, the Snake Monster can actively and passively conform its footsteps to traverse unmodeled terrain with relative ease and minimal computation. Using the mechanical spring compliance built into each joint, this robot conforms to uneven terrain rather than explicitly planning and optimizing footholds and step locations.

So far, the snake monster had no vision system, and used compliance to understand the environment. Compliance aims towards either process improvement (active) or human safety (passive), more about compliance can be found here [2]. During a compliant movement, the robot interacts with the environment and changes its behavior depending on whatever forces it senses. Tasks such as stair climbing, crawling, obstacle avoidance add much more complexities in coming up with a compliance algorithm. Even though we come up with a suitable algorithm, it will be having many constraints that need to be kept in mind. Adding a vision system to the current mechanical structure, eases many of the constraints. Hence, we come up with a vision system and some methods that can be implemented using such a system to improve the robot's capabilities. We demonstrate that the vision system is useful for better perceiving the environment compared to simple compliance system.

II. Modeling

A. *Mechanical and system architecture*

The six legs have a reach of 12 inches (30 cm), and are connected to a rectangular body, with the whole robot weighing 18 pounds (8 kg). The robot can move with an alternating tripod gait, with three legs in the air always. The robot can also operate in quadruped configuration. The robot has series elastic actuators (SEA) placed at each of its joints. More information about SEA can be found in [3]. Fig 1-b shows a series elastic module that has been developed in the Biorobotics lab and used as a building block of the snake monster. The modules use ethernet communicates through ethernet. Each of these SEA modules are embedded with Inertial measuring unit (IMU) and a torque sensing unit. These provide a base level component for the compliance model.



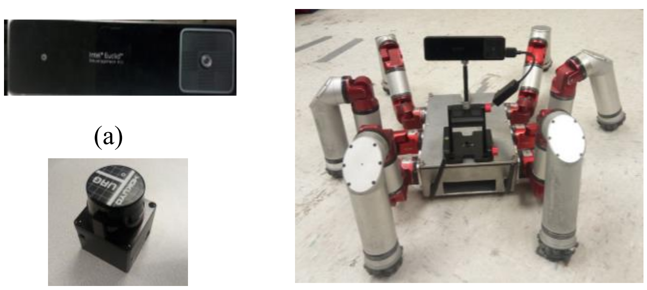
(a)

(b)

Fig. 1. (a) Snake Monster without a vision system, simply based on torque sensing (b) a series elastic module

B. *Vision Architecture*

The vision system of the robot consists of Intel Euclid, which is the integration of a depth camera and a motion camera. Along with an Intel-Atom™ x7-Z8700 Quad core CPU to produce an all-in-one compute and depth camera device in a compact and sleek package. We use this sensor to reduce the computation burden on the on-board processor. Unlike the traditional cameras which just pass the sensor data to the onboard computer, here the sensor itself is a unit of perception. Hence, the vision system perceives the environment and sends high level commands to the on-board computer. These high-level commands will be used for the functioning of the robot. We can observe the modularity in this system, which makes this model more versatile and reusable with similar sensors.



(a) (b) (c)

Fig. 2. (a) Intel Euclid camera (b) Hokuyo URG Laser, 2D-LIDAR (c) Then Snake Monster along with the vision system mounted on it.

III. METHODS

In this section, we list the various vision algorithms and tools that we tested and used on the Snake Monster to allow it to perceive its environment. We will talk about a SLAM algorithm (LSD slam), a mapping algorithm (Hector mapping), and a tracking algorithm (person recognition and tracking)

A. *LSD SLAM*

In this section, we briefly describe the LSD (Large Scale Direct) algorithm. For more details, the reader is encouraged to refer to the original LSD paper [4]. LSD is a direct featureless monocular SLAM algorithm that allows to build large-scale consistent maps of the environment. The algorithm uses highly accurate pose estimation based on direct image alignment, and reconstructs the 3D environment in real-time as a pose graph of keyframes. The global map is represented as a pose graph consisting of keyframes as vertices with 3D similarity transforms as edges, and this allows for incorporating multi scale environment nature. This algorithm runs in real-time on a CPU.

The algorithm consists of three main parts which are tracking, depth estimation, and map optimization.

For tracking, the rigid body pose is estimated with respect to the current keyframe. This part continuously runs at the highest frequency of the pipeline. The depth map estimation part basically uses the tracked frames from part one to either replace the current frame if some conditions are met, or to refine this current frame. The third part which is map optimization consists of including the keyframe that has been replaced in the previous part (depth map estimation), and thus means that it is not going to be refined anymore, into the global map. This parts also contains loop closure and scale drift detection which is done by estimating a similarity transform using a method called direct-sim(3)-image alignment (more details in the original paper [4]).

For initialization, the first keyframe is initialized with random depth map and large variance. LSD best initializes if a sufficient translational camera movement is detected in the first few seconds.

As of most state of the art SLAM algorithms, the map is represented as a pose graph of keyframes that captures the dependencies between all the keyframes that are added to the map. Each keyframe consists basically of an image, an inverse depth map, and variance of the inverse depth map. The edges of the pose graph represent similarity transforms between the keyframes.

B. *Hector mapping*

Hector mapping is a SLAM approach which does not use the odometric information, it works well on platforms which do not exhibit much of roll, pitch and yaw. It leverages the high update rate of the Laser to construct the occupancy grids. It constructs good resolution occupancy grids for environments which are smaller in area and have a lot of boundaries. The draw backs of this system are it cannot account for loop closure. It is sufficiently accurate for many real-world scenarios, more can be found in[5].

C. *Person recognition and tracking*

In this section, we describe our implemented method of person recognition and tracking. Person tracking enables the robot to recognize and follow a specific person. This gives the robot an ability to differentiate to obey the commands of this person, and enables it to differentiate this person from other people. We fuse the person tracking and color detection algorithms to track a specific person among a crowd of people. We provide the information about this person’s face and the color of the costume that he is wearing. First the robot detects a person’s face and after detecting the face it uses a bounding box to track the person. We extract the geometric center of this bounding box by computing centroid by taking the mean over the box and use that to stabilize the position of the Snake Monster. After tracking a person’s face and constructing a bounding box, the robot tries to match the color of the person’s costume, with the current color in the bounding box. If both the color and a face are detected, then the robot tracks this person, otherwise, it stays in its place.

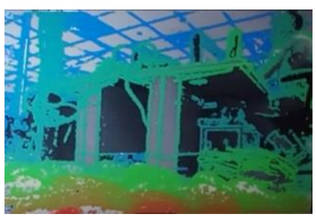
IV. Results

In this section, we describe the results that we got from applying the above-mentioned vision algorithms on different data sets collected using the Snake Monster.

A. LSD SLAM

Using the monocular sequence of images generated by the on board monocular camera, pointcloud of a rocky terrain was generated. The reason for using a monocular camera was based on an effort to make the vision system cost effective. The LSD algorithm was tuned such that it stays invariant to high camera noise. The camera captures images at 30Hz. The information was recorded in a bag file and played back at half the recording speed. The results shown in Fig. 2. (a) are the results for offline pointcloud computation. The reason for choosing offline computation was due to the high computation time demand, by LSD algorithm. As the size of the map increases with time, the algorithm starts slowing down. As a result, the time taken to construct the newer regions in the map increases. This delay in the map construction sometimes gives you an error of dropping frames during to the computational overburden. This leads to either less informative regions or corrupted maps (maps with too many) outliers.

(a)



(b)

(c)

Fig. 3. (a) LSD generated pointcloud (b) Disparity map of the current scene with respect to previous scene (c) The real image of the rocky terrain.

B. Person Tracking

We modified the person tracking algorithm given written by Intel so that it tracks only people with yellow costumes. As you see in the following figure, the algorithm correctly tracks the person who is wearing a yellow coat but doesn't track the same person if he is earing another color. For a demonstration, we made the robot follow a person who wears a yellow coat.

(a)

(b)

Fig. 5. (a) Person with yellow coat is tracked (b) Same person without coat not tracked

V. Future Work

What we have included in this paper so far is basically a bunch of vision based algorithms that enables any robot of perceiving its environment, and performing intelligent tasks that wouldn't have been possible otherwise (person tracking). Yet, a lot of additions and improvements can be done on the Snake Monster platform in the presence of all the sensor that we mentioned previously (Intel Euclid, 2D lidar, monocular camera, IMU). One extension of our work is basically combining the IMU readings with the 2D map built by the 2D-lidar, which will allow us to avoid drift and allows for loop closures. Another possible extension of our work is using the Euclid to reconstruct the terrain in front of the Snake Monster, thus allowing it to avoid obstacles during its navigation. Furthermore, we are planning to enhance the intelligence of the Snake Monster by allowing it to recognize hand gestures of a person, and doing actions accordingly.

Acknowledgment

I humbly acknowledge Hadi Salman, Puneet Singhal and Matthew Travers for their support in the project.

References

[1] https://en.wikipedia.org/wiki/Sago_Mine_disaster

[2] A. Albu-Schaffer et al., "Soft robotics," in IEEE Robotics & Automation Magazine, vol. 15, no. 3, pp. 20-30, September 2008. G. A. Pratt and M. M. Williamson, "Series elastic actuators," Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots, Pittsburgh, PA, 1995, pp. 399-406 vol.1.Ref 3.

[3] Engel J., Schps T., Cremers D. (2014) LSDSLAM: Large-Scale Direct Monocular SLAM. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision ECCV 2014. ECCV 2014.

[4] Stefan Kohlbrecher, Johannes Meyer, Thorsten Graber, Karen Petersen, Oskar von Stryk, and Uwe Klingauf, "Hector Open Source Modules for Autonomous Mapping and Navigation with Rescue Robots".