Design of a Fully Vision-Guided Mobile Robot

Masaharu Kobashi
University of Washington
Department of Computer Science and Engineering
Seattle, Washington 98195
mkbsh@cs.washington.edu

Akifumi Kobashi and **Atsuhide Kobashi**
Henry M. Gunn High School
Palo Alto, California 94306

Abstract

Most current mobile robots rely on laser or sonar range finders wholly or partially. We have designed a mobile robot which uses only visual information for navigation. Our robot design is based on our fundamental philosophy that high performance vision requires very high computing power. In order to fulfill this requirement, our robot is built to accommodate up to five high performance computers. Another unique characteristics of our robot is its visual SLAM and the cooperative 3D perception that uses both depth from focus/defocus and 3D binocular stereo.

Motivation

Currently most mobile robots rely on non-visual range finders (typically laser or sonar based) to interpret the 3D structure of the environment. The range-finder-based interpretation of the environment is simpler to implement than vision-based systems. However, range finders capture very little properties of the real environment. They cannot get rich visual information that leads to the ultimate goal of human-like perception of the real environment. Range finders are also limited by their effective measurable ranges, which depend on the type and the model of the range finder.

Our robot is a concrete instance of our attempt to overcome the limitations of range-finder-based perception. It uses vision to interpret the environment and guide the robot. In order to achieve this goal, the robotic platform has to be capable of accommodating powerful computers, since vision, which is implemented by hundreds of millions of parallel units in the human eyes and brain, requires enormous computing power. It translates to the necessity for the physical capability to carry plenty of batteries to run powerful multiple computers. To meet these requirements our robot has a very sturdy structure and powerful motors that are capable of carrying hundreds of pounds of equipment and power sources. Another unique aspect of our robot is in the fully computer-controllable video cameras and the camera mount. They are designed for sophisticated active vision algorithms for reliable 3D perception.

Figure 1: Our robot

Hardware Design

We did not buy a ready-made robot, because we could not find on the market any robot that satisfied all of our requirements for the fully vision guided robot. Instead, we designed the robot from scratch and built the whole body from aluminum bars and sheets fabricated by our machine shop. The following are the descriptions of the major components of the robot.

Cameras

We installed two CCD-based IEEE1394 (firewire) video cameras whose focus (lens position) and aperture can be controlled from a computer using the IIDC protocols. The adoption of the IEEE1394 and the IIDC protocols allows us to use various publicly available open-source libraries.

An important feature of our robot is its use of cameras with computer-controllable focus and aperture settings. Most robots do not have this capability. They use cameras with a deep depth of field, so that the need for focus adjustment is not significant. We need this capability, because our robot performs the 3D measurement by both depth from focus/defocus and the binocular stereo. By using two different methods for 3D perception, our robot can achieve a better 3D interpretation. The focus/defocus information supplements the weakness of the binocular stereo system and vise-versa.

Camera Mount

The two video cameras are mounted on a platform whose pan, tilt, and vergence are controllable from a computer. Four servo motors are used for the camera mount, two for vergence control (one for each camera), one for pan, one for tilt. The pan range is 270 degrees and each of the upward and downward tilt ranges is 45 degrees.

Computers

The robot can accommodate up to five full-size ATX motherboards. They are to be used for distributed computing to cope with the demand for the real-time high computing power especially needed for vision and visual SLAM. As of this writing, we have not used the full capacity of the robot with respect to the number of computers it can accommodate. As a first experimental step, we used two computers, one for the visual processing, the other as the master controller which handles all the other tasks including path planning and controlling the drive motors and the four servo motors on the camera platform.

Power Source

We chose 12 volt AGM deep cycle lead-acid batteries for the robot's power source. We selected lead-acid batteries because of their economy and time-tested stability, although their capacity/weight ratio is not ideal.

The robot uses 24 volt DC power for both the driving motors and the computers. The 24 volt DC power is created by connecting two 12 volt batteries in series. For the driving motors, the 24volt DC power goes to the motors by way of a high-current motor controller. For the computers the 24 volt DC power goes to a DC-AC inverter to supply the AC power to the computers whose power supply modules are standard AC-DC type.

Although it is simpler and may seem more efficient to use DC-DC conversion for computers instead of going through the lengthy DC-AC-DC (computers need DC) conversion, we did not choose the DC-DC for the following two reasons. The total efficiency of the DC-DC and the DC-AC-DC are almost the same with available converters within our budget. Moreover, with DC-AC-DC we can use our existing computers without replacing their power supply modules with new DC-DC type power supply modules.

Some of the conditions above will no longer be valid in the near future. For example, with the reduction of the price of the lithium-ion batteries and possible further improvement of their capacity/weight ratio, we may switch to lithium-ion batteries in the near future. We may also switch the power supply architecture for computers from DC-AC-DC to DC-to-DC in the future, if efficient high-power DC-DC converters for computer motherboards become available at reasonable prices.

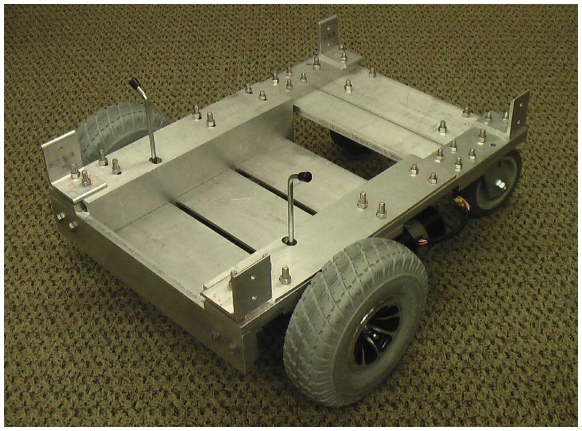

Figure 2: Chassis of the robot

Drive Train

The robot is driven by two 500 watt gear-motors, each of which drives one of the front wheels individually. These motors can carry the robot at the maximum speed of 4 miles per hour with the full load (350 pounds) . We chose two strong gear-motors in order to carry the huge amount of batteries that are needed to run the multiple high performance computers. The rotation of the gear-motors can be reversed so that the robot can go both forward and backward as well as turning in all directions while standing. The smallest turning radius of the robot is 19 inches.

Motor Controllers

We selected a high power dual channel DC motor controller that is capable of handling up to 120 amp of total current (60 amp per channel). We wrote drivers for this controller by ourselves, so that high-level commands of our choice can be used in the motor control module.

For the four servo motors which are used for the camera mount, we used two PWM servo controllers. One of them is used for the two servo motors on the camera platform to control the vergence and the other for the two servos in charge of pan and tilt. Each of the three controllers is connected to the serial port of a computer using the RS232 protocol.

Physical Structure

Our robot is designed to be capable of carrying multiple high performance computers to enable sophisticated vision processes. Its chassis must be particularly sturdy to handle the weight of over 400 pounds, as the total weight of batteries alone can be 250 pounds. The batteries are placed at the bottom on the floor, which is inset to the level of the front wheel axle in order to have a low gravity center. The low gravity

Dimension	26.5”(width) 29”(length) 43”(height)
Body Material	Aluminum 6061
Drive Motors	Two 500 Watt gear-motors
Power Source	Up to 250 pounds of AGM lead-acid batteries for both drive motors and computers. A high power inverter is installed for DC-AC conversion for the computers.
Computers	Accommodates up to five full size ATX motherboards
Cameras	Two IEEE1394(Firewire) video cameras capable of controlling focus and aperture from a computer
Camera Mount	Pan-tilt-vergence controllable mount with four servo motors
Remote Control	Controllable wireless by the protocol 802.11g for training and emergency
Wheels	Two 10 inch diameter front wheels and two 6 inch diameter rear swivel wheels
Turning Radius	19 inches (Minimum)
Speed	4 mph (Maximum) at full load

Table 1: Hardware Specification

center is needed, since the robot is designed for both indoor and outdoor use.

The robot's dimension is also designed for both indoor and outdoor use. The width and the length are limited so that the robot can go through any standard door. For the structural material we chose aluminum 6061 for its good balance of strength, machinability and cost.

Software Design

We selected Linux for the robot's operating system. Multiple computers are used for performing CPU-intensive tasks in real time using a distributed computing architecture. Major tasks performed by the computers are the 3D depth measurement, visual SLAM, path planning, and motor control.

Multiple computers are connected for distributed computing. As of this writing, we connected them by the standard server-client architecture, although we plan to change the design in the future as we install more high-performance computers. The distributed computing will be based on MPI when the robot takes advantage of its full capacity.

3D perception

One of the unique characteristics of our robot is its use of both depth from focus/defocus (Pentland 1987; Nayar, Watanabe, & Noguchi 1995) and binocular stereo to compute the 3D depth. Both binocular stereo and depth from focus/defocus have strengths and weaknesses. Our attempt is to take advantage of the strengths of both and to avoid their weaknesses.

Binocular stereo is generally capable of measuring the 3D depth more accurately than depth from focus/defocus methods, because the base-line of the binocular setup is usually much longer than the diameter of the object lenses of the cameras. However, binocular stereo fails to measure the 3D depth when correspondence cannot be established correctly between two images from the two cameras. Types of image regions with which correspondence fails depend on the feature detector. For example, SIFT (Lowe 2004), probably the most used feature detector, does not detect long straight lines by design, although they are very conspicuous to the human eye. It also fails if a critical part of a region, even if very small, is detected in one camera's image and is slightly occluded in the other.

On the other hand the focus/defocus method can measure the depth in some cases where binocular stereo fails. For example, regardless of type of shape, (straight line or corner), the focus/defocus method can sense the depth. For another example, a region having multiple identical or very similar patterns close to each other (e.g. regularly aligned dots or grids), can often derail the corresponding algorithms. Again depth from focus/defocus can correctly measure the depth for such regions. For this complementary relation between the two methods, we use both of them to enhance the robustness of the depth measurement in unrestricted natural environment.

Another unique aspect of our 3D measurement is the feature detection step. For our binocular stereo algorithm, we do not use any of the well-known feature detectors such as SIFT (Lowe 2004), which is probably the most used feature detector for correspondence. There are a number of other well-known alternatives such as Harris Affine (Mikolajczyk & Schmid 2004), MSER (Matas *et al.* 2002), and there is a detailed performance comparison (Mikolajczyk *et al.* 2005) of these feature detectors. However, our robot uses a conspicuous-region-based feature detector we developed. Our detector performs robustly in the real environment and the detected regions are closer to the regions conspicuous to the human eye than other well-known detectors.

Our robot also takes advantage of the active use of vergence control in order to enhance the reliability of the 3D depth measurement. Measured distances are used for further processes with probability of accuracy. The further the measured distance, the less accurate is its measurement. Beyond a certain distance, the robot does not try to compute accurate measurements. They are given a special “very far” mark, as such distant objects are irrelevant to the maneuver of the robot.

Visual SLAM

SLAM systems usually uses only range data, which is obtained from laser or sonar range finders. Our robot uses visual SLAM (Se, Lowe, & Little 2005; Gil *et al.* 2006) which does not use a non-visual range finders (laser or sonar based range finders). It analyses the visual data from the cameras for building/updating the map of the surrounding region and localizing itself. Our visual SLAM has the following characteristics.

- While the standard SLAM uses depth information calculated only at single points, our visual SLAM takes advantage of rich image information including shapes, color,

and gradient. The rich visual information makes it possible to identify parts of scenes more accurately than by simple range data of points.

- Our system uses topological information, which is robust and invariant against noise and inaccuracies in measurement under even substantial viewpoint changes.
- Our robot frequently looks around and looks back to check and update the scene properties at various angles and distances. Recording scene properties from multiple views enhances the robustness of the map and it is especially effective to find a return path.
- A disadvantage of our visual SLAM is it needs more computing power and more data storage space than the standard SLAM, since visual information is more complex than simple range data.

Execution Flow

At the beginning of a mission, the robot is given a goal. First, the robot calculates its current pose and location and computes a path to reach the goal. For the path planning we adopted a strategy that computes a path that is made of multiple connected sub-paths each of which is either a straight line or an arc.

As shown in Figure 3, at the start position of each sub-path, the robot regards the end of the sub-path as an intermediate goal and generates a sequence of low-level motor commands that are needed to reach the intermediate goal. While executing the motor commands, the robot constantly observes the ego-motion with respect to the surrounding scene visually and, if necessary, corrects the power of each motor. At the end of the sub-path, the robot updates the map as well as its pose and location. The updated data of the map and the robot's pose and location are used to update the path plan. The end of the next sub-path in the path plan is set as the next intermediate goal, and this loop continues until the robot reaches the goal.

Although the step labeled "Execute Commands with Adjustment by Vision" in Figure 3 appears simple, it actually is a complex process, since we did not use even encoders. However, our effort to use vision only even in that step will pay off when we run the robot on a rough or slippery ground where encoder reading is not completely reliable.

Performance

As of this writing, we have not taken advantage of the full capacity of the robot with respect to computing power. As stated before, our robot is capable of carrying five computers. However, our experiment has been done using only two medium-speed Pentium 4 processor-based computers. Because of this preliminary setup, our robot has to stop frequently to analyse the scene. Another problem that our robot occasionally encounters is it makes wrong decisions on the floor patterns. With some patterns on the floor, our robot fails to make distinction between flat floor and obstacles. However, generally it performs well in obstacle avoidance and path following in the indoor situation. Testing in the outdoor environment has not been done as of this writing.

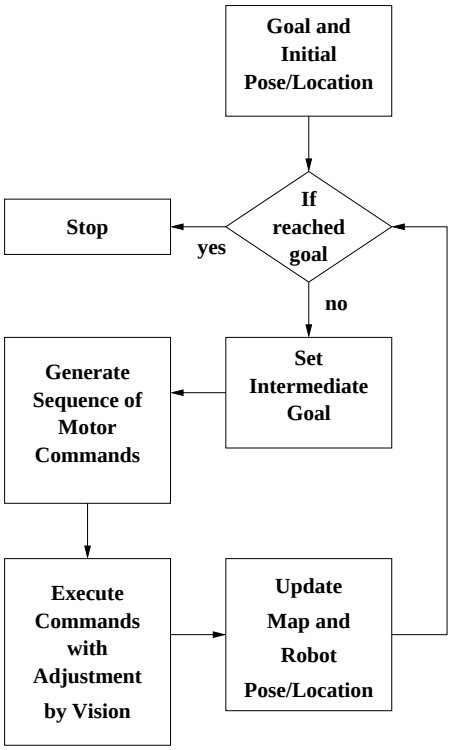

Figure 3: Execution Flow

Future Work

The first agenda for improving our robot is to increase the computing power by installing at least three, hopefully five, high-performance computers. Currently, we are using a standard client-server architecture for distributed computation, as the moving speed of the robot is restricted and we allow the robot to stop frequently for decision making. However, as we increase the number of computers, we will deploy more sophisticated distributed computing architecture using MPI. We can enhance the computing power further by changing the power conversion method from DC-AC-DC to DC-DC and remove the DC-AC inverter. In this way we can use the shelf space of the inverter for installing another computer, making the total 6 computers.

Second our 3D perception and visual SLAM are still under development. Currently we are working on developing an improved version of our feature detector and more robust and efficient matching algorithms. Finally, we will install two arms taking advantage of the sturdy body design and the focus adjustable cameras. The installation of the arms does not require a major surgery, since our robot was designed from the beginning to accommodate two arms with substantial holding power. The focus adjustable cameras enable the robot to watch both the close objects to manipulate and the far objects relevant to path planning. We hope to use our robot for two vision related research: vision-guided robot maneuvering and visual servoing for manipulating objects.

References

Gil, A.; Reinoso, O.; Burgard, W.; Stachniss, C.; and Martínez Mozos, O. 2006. Improving data association in rao-blackwellized visual slam. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2076–2081.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2):91–110.

Matas, J.; Chum, O.; Urban, M.; and Pajdla, T. 2002. Robust wide baseline stereo from maximally stable extremal regions. In Rosin, P. L., and Marshall, A. D., eds., *BMVC*. British Machine Vision Association.

Mikolajczyk, K., and Schmid, C. 2004. Scale & affine invariant interest point detectors. *International Journal of Computer Vision* 60(1):63–86.

Mikolajczyk, K.; Tuytelaars, T.; Schmid, C.; Zisserman, A.; Matas, J.; Schaffalitzky, F.; Kadir, T.; and Gool, L. V. 2005. A comparison of affine region detectors. *International Journal of Computer Vision* 65(1/2):43–72.

Nayar, S.; Watanabe, M.; and Noguchi, M. 1995. Real-time focus range sensor.

Pentland, A. P. 1987. A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.* 9(4):523–531.

Se, S.; Lowe, D. G.; and Little, J. J. 2005. Vision-Based Global Localization and Mapping for Mobile Robots. *IEEE Transactions on Robotics* 21(3):364–375.