Fourier Techniques and Monocular Vision for Simplistic and Low-Cost Visual Odometry in Mobile Robots.

Prepared by:
Ricardo Ramirez

Faculty Advisors:
Dr. Randy Hoover

Research Experience for Undergraduates
Summer 2016

Site Director:
Dr. Thomas Montoya

South Dakota School of Mines and Technology
501 E Saint Joseph Street
Rapid City, SD 57701

Table of Contents

Abstract ………………………………………………………………………...3
Introduction ……………………………………………………………….…3
  Motivation …………………………………………………….…..4
  Objective ………………………………………………………..4
Procedure ………………………………………………………………….5
  Equipment ……………………………………………………...5
  Method ……………………………………………………………..6
  Experimental Procedures ……...……………………………………...7
Results …………………………………………………………………….8
Discussion ………………………………………………………...…9
Conclusion ………………………………………………………………..10
References ………………………………………………………………..11
Acknowledgements ……………………………………………………………11

Abstract

Despite their low cost, webcams offer impressive resolution, with low-end webcams offering resolutions of up 1280 x 720. This gives them potential of being low-cost, yet efficient vision sensors for mobile robots. Specifically, a webcam can be used as a sensor for visual odometry, where a mobile robot localizes its position (and orientation) as it moves through an environment. This investigation presents a technique for low-cost mobile robot localization using a single downward-facing camera (monocular vision). Utilizing an Frequency based image registration technique, translation distance and rotation angle between consecutive video frames can be determined via normalized cross correlation. This results in relative position and heading estimation of the robot using only webcam images. Results of this technique are performed through navigation experiments on a robot operating in a controlled environment. It is shown that this technique has the ability to aid in the localization process providing a low-cost sensing solution.

Introduction

Localization and mapping are important capabilities for autonomous mobile robots in order to navigate intelligently in an environment. Localization refers to as knowing where the robot is within an environment and mapping as using that knowledge to develop a map of where it has traveled. Localization and mapping are typically performed using a combination of sensors, including wheel and visual sensors (cameras) [3].

Depending on the task, autonomous mobile robots typically take information from all of these sensors combined to accurately estimate robot position and heading from its motion over time [1]. This estimation is typically referred to as the robot's odometry . When this estimation involves the use of images from a visual sensor, it is called visual odometry.

Visual odometry (VO) offers a natural complement to the other sensors a robot might use for localization purposes. For example, when relying on wheel encoder information in sand and gravel surfaces, levels of uncertainty arise due to variable wheel slippage on these surfaces. In these situations, visual odometry has shown improvement over odometry from wheel encoders [1], [3].

Motivation

As smaller, less expensive mobile robots become more prevalent, new challenges in terms of sensing accuracy may arise. Strict cost or size constraints may be imposed on overall design. In such cases, the use of multiple, sophisticated and high cost sensors may not be a feasible option. Robots for exploration purposes, for instance, may incur damages or fail, which makes cost a major issue in their deployment. One way to deal with this issue is to rely on less sensing equipment. However, having less sensing equipment typically leads to a decrease in accuracy of localization. Clearly, a tradeoff between cost and accuracy arises in attempting to achieve accurate and reliable localization. The work presented in this paper is motivated by this tradeoff. It is an exploration on attaining comparable localization accuracy by imposing a strict constraint of relying on only one sensor, namely a low cost camera.

Objective

Simple and low cost cameras such as webcams offer fair resolutions and video capabilities for their price. This gives them potential of being low-cost, yet efficient vision sensors for mobile robots. Past work has shown that robots operating close to the ground, mounted with a downward-facing camera, can take advantage of commercial webcams to facilitate in the localization process. This camera-robot configuration achieved success in localization when used with a Fourier transform technique of image registration [1], [3]. This paper presents work that is similar to [1], which similarly uses a webcam pointed at the ground to aid in the localization process.

Procedure

Equipment

Low cost and simplicity was a major focus of this study. Therefore, a single webcam was the only sensor mounted on the robot, thereby making it a monocular rather than stereo VO system. During performance trials, two different cameras were tested to compare their relative advantages and disadvantages.

The iRobot Create—a reprogrammable robot for hobbyists, researchers, etc.—was chosen in this study, also for its simplicity and low cost. To fasten the webcam, the top of the robot was first fitted with a thin wooden platform. The webcam was then fastened to the platform using an approximately 8 inch aluminum “boom arm,” which extended the camera away from the robot approximately 3 inches.

iRobot Create

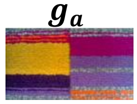
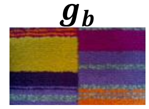
Cameras tested for VO system

Method

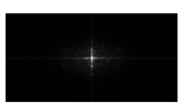
In order to arrive at the ultimate goal of robot localization, images captured during motion must first be registered using one of several methods. This study used a Fourier Transform (FFT), which transformed captured images into the frequency domain. To estimate relative translation and rotation between subsequent video frames a phase correlation approach was used. The method was programed in MATLAB and made use of the Image Acquisition Toolbox. The following demonstrates the usage of phase correlation to determine relative translative movement between subsequent two video frames.

Phase Correlation Method:

1. Given 2 images $\boldsymbol{g_a}$ *and* $\boldsymbol{g_b}$

$\boldsymbol{g_a}$	$\boldsymbol{g_b}$
	

2. Calculate the discrete 2D Fourier Transform of both images: $\boldsymbol{G_a = F\{g_a\}, G_b = F\{g_b\}}$

Magnitude Spectra of First Frame	Phase Spectra of First Frame
	
Magnitude Spectra of Second Frame	Phase Spectra of Second Frame
	

3. Calculate the cross-power spectrum by multiplying the first Fourier transform and the complex conjugate of the second and normalizing the product elementwise [2].

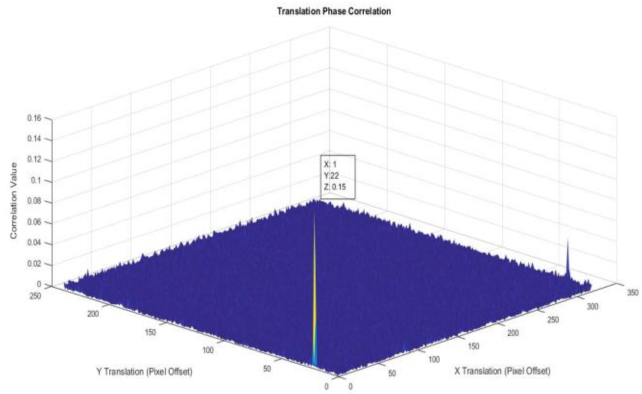
$$\boldsymbol{R = \frac{G_a \circ G_b^*}{|G_a \circ G_b^*|}}$$

4. Apply the inverse Fourier transform to obtain the normalized cross-correlation.

$$\boldsymbol{r = F^{-1}\{R\}}$$

5. Determine the location of the peak in **r.**

$$(\Delta x, \Delta y) = argmax_{(x,y)}\{r\}$$



Experimental Procedures

Fastened with the camera, the robot captured images while driving across one of two types of surfaces. Surface A is a bare floor in a laboratory. Surface B is a carpeted floor. As seen below, these surfaces are very different. One is nearly bare and featureless while the other is highly patterned. It was believed that video should be captured on surfaces that were representative of either extreme—low feature content vs high feature content. The phase correlation method above was exploited to compute total displacement of the robot as it captured frames while driving forward or rotating over either one of the surfaces.

Surface A **Surface B**


It was believed that to ultimately arrive at localization, it was first necessary to verify through exhaustive trials that the phase correlation method was successful in consistently and accurately computing total displacement of the robot while it drove forward or turned in place over either surface. To that end, two sets of displacement trials were performed—driving straight forward and in-place rotation.

Throughout the trials camera resolution and video frame rate of both cameras as well as robot speed were varied and monitored—resolution was switched from 320x240 to 640x480 pixels. Frame rate was varied between 15fps to 30 fps. In addition various robot driving speeds between 5 cm/sec to 100 cm/sec were tested. After each trial total VO displacement distance was compared to the robot's internal wheel odometer to check accuracy. The total linear displacement of the robot was then computed when the robot finished capturing video frames and stopped forward motion. This displacement calculation was computed in less than a second from when the robot stopped.

Results

Driving Forward

The first set of trials were performed on Surface A. Exhaustive trials showed that our VO approach failed to accurately and consistently compute linear displacements even after a thorough exploration of camera settings and robot speeds. In short, two typical problems of image registration proved difficult to overcome, namely, specular reflection and shadowing.

The second set of trials were performed on surface B. On this surface specular reflection was eliminated and shadowing was minimized since the colors of the carpet were dark. Throughout these trials, motion estimation reached a modest level of accuracy. Through most trials, and with the varying camera resolutions, the total displacement of the robot measured by the wheel odometer and our VO generally differed by 8%. This percentage difference sustained through exhaustive trials.

Rotation in Place:

All of the rotation in place trials were performed on Surface B. The robot was commanded to turn 90° (either left or right) and the total pixel displacement in both the X and Y direction was recorded. Unfortunately, through exhaustive trials the total X and Y displacement never remained consistent and differed by unreasonable amounts across trials. The various camera settings were exploited and rotation speed was varied throughout, yet no consistency of total X and Y pixel displacement was achieved.

Discussion

Before trial phase began, the accuracy of the wheel odometer was tested by taking robot displacement measurements with a yardstick over half meter and one meter lengths. Multiple measurements revealed that the wheel odometer was not consistently accurate. Generally, the difference in wheel odometry and physical measurement differed by as much as ±3 cm across multiple measurement attempts.

As seen from the results of calculating forward translation, this low-cost approach to VO showed potential in calculation linear, i.e., driving forward. The approximately 8% difference that subsisted throughout the trial period was possibly due to an inaccuracy in cm/pixel estimation that was measured in the preliminary stage of setup. The obvious limitation is, however, that the robot must be on a surface with high feature content such as a patterned carpet to attain these results.

With respect to rotation, the outcome of the performance trials showed much more work needs to be done. Evidently this approach to VO failed to reach any level of consistency in computing total pixel displacement in left or right rotations. We believe, however, that with more work, it is possible to determine how to overcome failure with rotational translation, given that modest success was achieved in linear translation. Therefore, this method should not be entirely discarded. The challenge of where, why, and how it fails is worth exploring for this unique approach. Until this is determined, this approach to VO is stalled. It is worth noting that this seemingly relatively simple approach reaches somewhat more complexity due to the position of the camera. Since it is placed away from the robot, and not dead center, translation is not planar like in past successful VO systems.

Conclusion

In sophisticated autonomous robots, visual odometry is used with wheel odometry and information from other sensors to give autonomous mobile robots robust localization capabilities. Visual odometry succeeds where other forms may have problems (variable wheel slippage in wheel odometry) and may be implemented with a low cost sensor such as a camera with fair resolution. In some autonomous robot systems, a camera may be the ideal and only sensor. Attaining accurate localization capabilities from only a camera is a task worth exploring since cost and/or size may be a constraint in robot design. When a single camera is used, an FFT approach to image registration to estimate robot motion has shown comparable localization results. A variant of this method was explored and was the work shown in this paper. After repeated experimental trials, the method proved promising results in computing straight forward displacement of the robot. When trial phase were done to compute rotation, the method looked less promising and required rigorous exploration due to the difficulties associated with calculating non-planar rotation.

References

[1] C. M. Gifford, "Low-Cost Mobile Robot Localization Using Only a Downward-Facing Webcam", Technical Report, University of Kansas, USA, 2009.

[2] B. Reddy, B. Chatterji, An FFT-Based Technique for Translation, Rotation and Scale-Invariant Image Registration, IEEE Transactions on Image Processing 5 (8) (1996) 1266–1271.

[3] Killpack, Marc, et al. "Visual odometry and control for an omnidirectional mobile robot with a downward-facing camera." *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010.

Acknowledgements

The funding for this research came from the National Science Foundation. Thanks to advisor Dr. Randy Hoover and REU site director Dr. Thomas Montoya for their direction and guidance.