NASA/TM—2012-216043



Localization Using Visual Odometry and a Single Downward-Pointing Camera

Aaron J. Swank
Glenn Research Center, Cleveland, Ohio

September 2012

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI Program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*

- E-mail your question to *help@sti.nasa.gov*

- Fax your question to the NASA STI Information Desk at 443–757–5803

- Phone the NASA STI Information Desk at 443–757–5802

- Write to:
  STI Information Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076–1320

NASA/TM—2012-216043



Localization Using Visual Odometry and a Single Downward-Pointing Camera

Aaron J. Swank
Glenn Research Center, Cleveland, Ohio

National Aeronautics and
Space Administration

Glenn Research Center
Cleveland, Ohio 44135

September 2012

Trade names and trademarks are used in this report for identification only. Their usage does not constitute an official endorsement, either expressed or implied, by the National Aeronautics and Space Administration.

Level of Review: This material has been technically reviewed by technical management.

Available from

NASA Center for Aerospace Information
7115 Standard Drive
Hanover, MD 21076–1320

National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Available electronically at http://www.sti.nasa.gov

Localization Using Visual Odometry and a Single Downward-Pointing Camera

Aaron J. Swank
National Aeronautics and Space Administration
Glenn Research Center
Cleveland, Ohio 44135

Abstract

Stereo imaging is a technique commonly employed for vision-based navigation. For such applications, two images are acquired from different vantage points and then compared using transformations to extract depth information. The technique is commonly used in robotics for obstacle avoidance or for Simultaneous Localization And Mapping, (SLAM). Yet, the process requires a number of image processing steps and therefore tends to be CPU-intensive, which limits the real-time data rate and use in power-limited applications. Evaluated here is a technique where a monocular camera is used for vision-based odometry. In this work, an optical flow technique with feature recognition is performed to generate odometry measurements. The visual odometry sensor measurements are intended to be used as control inputs or measurements in a sensor fusion algorithm using low-cost MEMS based inertial sensors to provide improved localization information. Presented here are visual odometry results which demonstrate the challenges associated with using ground-pointing cameras for visual odometry. The focus is for rover-based robotic applications for localization within GPS-denied environments.

1 Introduction

Conventional navigation techniques such as GPS do not presently offer adequate position knowledge inside of buildings or below the earth's surface. Numerous challenges exist for utilizing GPS in these environments, including weak or lack of signal, dilution of precision and multipath effects. In addition, for robotic or human surface exploration beyond the Earth, the GPS infrastructure is not available. The localization of an individual or robotic platform is often desired for these environments. Numerous approaches have been presented in the research community for navigation knowledge within GPS-denied environments. The use of vision-based sensors to provide navigation information is a common alternative to large and expensive inertial measurement devices. Vision-based sensors may also be used to augment inertial devices to reset errors due to inertial drift. These errors are especially apparent in non military-grade or inexpensive MEMS inertial sensor devices.

A common approach to visual navigation is the use of stereo imaging to replace the inertial navigation system. The Mars Exploration Rover, for example, extracts visual odometry information from a pair of stereo images [1], [2]. For stereo-based image processing, the required image pre-processing tends to be CPU-intensive and limits the real-time data rate. As an alternative, here we investigate the use of vision-based sensors to augment inexpensive MEMS-based navigation systems.

For wheel-based robotics, incorporation of wheel odometry measurements is one simple method for providing additional sensor input to the navigation estimation algorithm. Traditional odometry methods utilize rotary encoders to measure wheel rotations. Motion estimation using odometry techniques alone is unreliable as the measurements suffer from errors due to slippage which will accumulate over time. Furthermore, wheel odometry on a robotic platform using skid-steering technology is even more unreliable, as turning is achieved through wheel slippage. As an alternative to traditional odometry methods, a vision-based sensor may be used for odometry measurements. Such visual odometry techniques are not prone to errors associated with wheel slippage like traditional wheel sensor techniques. This work further investigates the use of vision-based sensors for odometry measurements.

2 Methodology

Visual odometry is performed by determining the position from sequential camera image analysis. In contrast to stereo-vision implementations, here only a single camera is used in order to reduce computational requirements by eliminating the requirement for feature-based stereo matching algorithms. The compromise is that with a single camera, a three-dimensional position is no longer computed for each selected

feature, and only two-dimensional translational information is obtained. The single downward pointing camera is intended to replace or augment encoder-based odometry for applications on skid-steering robotic platforms. The concept is akin to that of an optical mouse. A similar visual odometry concept with a single downward-pointing camera is discussed in Reference [3], but the approach and implementation is different. Here, the approach is to use feature tracking between two temporally spaced image frames to construct an optical flow field. The relative displacement between each image frame is then extracted and used as an input to a position estimation algorithm. The basic methodology follows closely that of a visual odometry system and includes:

1. **Frame Acquisition:** Acquire temporally spaced images.
2. **Feature Detection:** Determine features in the image frames to use for tracking.
3. **Optical Flow:** Use the tracked features to perform optical flow calculations between the two image frames.
4. **Sensor Filtering:** Filter noisy data using estimation algorithms. Eliminate erroneous data and obvious outliers.
5. **Localization:** Use optical flow measurements in motion estimation algorithms to determine position.

Typically, vision-based navigation systems also perform an image correction step to account for errors such as lens distortions. Here for simplicity reasons, no corrections are made to the image to account for these errors. Yet the general method allows for the corrections to be applied as needed. In addition, for the downward-pointing camera it is assumed that the image plane is sufficiently parallel to the surface ground plane such that a change in pose corresponds directly to a planar transformation of the images. Again, a correction step may be needed to satisfy this assumption.

3 Implementation

3.1 Testbed

For initial proof of concept demonstration purposes, a localization experiment using a visual odometry setup is constructed. The experimental setup consists of hardware currently available on-hand without any additional procurement. A crude "duct-tape integration" approach accurately describes the setup. All of the hardware is attached to a laboratory cart, which contains four swivel caster wheels. The caster wheels allow the cart to be moved in any direction within the plane of the floor. It should be noted, that a setup with rotating caster wheels is actually

a more challenging configuration for position and trajectory estimation than that of a rover, where the wheels or skids provide constraints to the direction of motion. With traditional wheels or skids, a change in the velocity to the cross-track direction requires slipping of the wheels or sliding of the tracks.

3.2 Frame Acquisition

For the visual odometry measurements a frame capture device is required. Although only a single video camera is used for the visual odometry measurements, two cameras are attached to the laboratory cart for performance comparison purposes. The first video camera is a common web camera. For this setup, a 640x480 image is acquired with the web camera. Images are acquired between five and ten frames per second, depending on the exact test configuration. The video camera is attached to a laptop computer where the frame capture data is available for near real-time visual odometry measurements.

The second camera attached to the laboratory cart is a high-definition (HD) video camera. The HD video camera (720p, 60 fps) is currently used only for comparison of results. The video stream from the HD video camera is recorded and then used for post-processing video odometry measurements. The available HD video camera is designed to dump the video data directly to a storage medium, and therefore is not available for real time processing. Both the web camera and the HD video camera produce color video. Since only gray-scale image frames are required for the visual odometry algorithm, color frames are converted to gray scale prior to use.

For repeatable visual odometry measurements and to ensure proper calibration, it is necessary to maintain a constant distance from the camera optics to the tracked object. For this application, the tracked features are on the ground. Maintaining a constant distance to the tracked object is not a restrictive requirement as the rover has wheels or tracks on the ground. Thus, by positioning the web camera to acquire images of the ground near the wheels, a constant distance to the imaged surface is maintained.

3.3 Feature Detection and Optical Flow

For feature detection and the optical flow analysis, the routines available in the Open Source Computer Vision library (OpenCV)[1] are used. Feature detection is implemented using Shi and Tomasi corner detection [4]. The optical flow analysis is then performed using a Pyramidal implementation of the Lucas-Kanade optical flow technique [5], [6]. The parameters used in the feature detection and optical flow tracking routines are found in Table 1.

[1]OpenCV is available from http://opencv.willowgarage.com

Parameter	Setting
Feature Detection:	
Number of Features	100
Quality Level	0.04
Minimum Distance	0.01
Block Size	3
Optical Flow:	
Window Size	3 x 3
Pyramid Level	5
Max Iterations	20

Table 1. Feature Detection and Optical Flow Parameters

One difficulty with performing optical flow measurements using temporally spaced image frames is the potential variation in illumination between images. Optical tracking algorithms assume a consistent illumination of the picture frame between successive frames. For this experimental setup the difficulty is overcome by adding a direct illumination source near the camera. The light source is placed at an angle to the observed surface such that there is no direct reflection back into the camera from shiny surfaces. In addition, by illuminating the imaged surface from a non-zero angle of incidence, shadows are produced from any textures, dirt, or surface roughness which may be present. The addition of these shadows aids the feature tracking routines by potentially generating more points of interest for feature tracking.

3.4 Sensor Data Filtering

When using visual measurements, the results often contain a number of outliers. Furthermore, methods which match "point" features are not usually robust. For this downward pointing camera approach, outliers in the results are especially likely. Pictures of the ground typically do not have ample features to provide transformation tracking. The images may also lack an adequate number of matching features between sequential frames to provide a reasonable statistical representation of the measurements. However, as will be demonstrated in this work, it is possible to achieve feature tracking using ground images. Figure 1 shows for example a representative optical flow measurement image frame using tracked features from a downward pointing camera. The image shows a number of flow lines with a common direction and magnitude. There are also a number of clearly erroneous tracked features depicted by the flow lines. By using proper estimation and data filtering methods, the optical flow data set can still be used to produce reasonable information. In practice, robust estimation techniques such as Random Sample Consensus

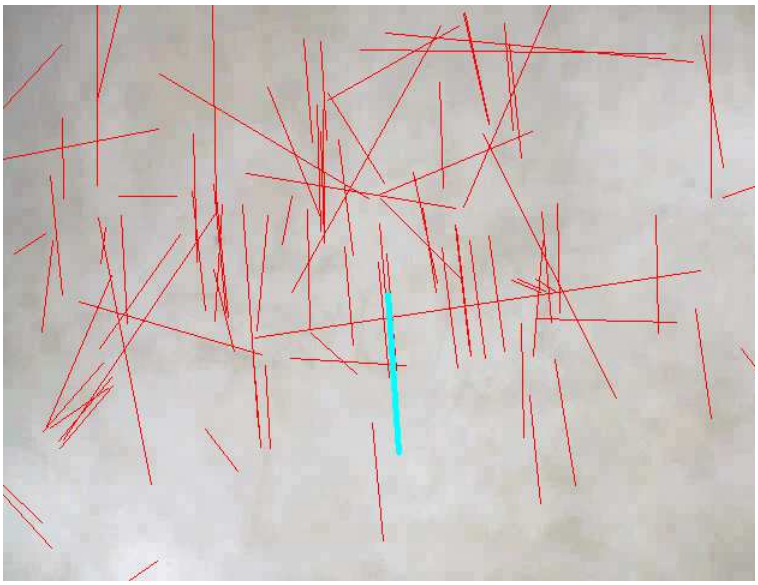


Figure 1. Optical flow results on a textured surface. The red lines depict offsets for tracked features. The bold cyan line in the middle indicates the estimated result of all the tracked features in the image.

(RANSAC) [7], or M-estimator Sample Consensus (MSAC) [8] are often used. One disadvantage of RANSAC-based methods, is that the routine is iterative and the number of iterations required between each successive estimation is not necessarily predictable. In addition, the thresholds set in the algorithm are likely to be problem specific. For these reasons, an alternative approach is taken here.

Close examination of the sample histogram for a number of optical flow results lends insight to the necessary filtering approach. The optical flow results may contain features that are not necessarily Gaussian. Furthermore, the characteristics between successive optical flow results may exhibit different characteristics. A collection of optical flow results were found to exhibit multimodal, heavy- or light-tailed characteristics as well as Gaussian distributions. Figure 2 shows for example a sample histogram for the direction of motion as estimated from an optical flow analysis. The histogram indicates a multimodal distribution is present in the data. In this particular instance, the correct result is associated with the mode of lower probability density.

For a proof-of-concept demonstration, it is desired to use standard techniques for which software libraries or toolboxes are easily obtained. For this work, the Kalman filter routine available in OpenCV is used. Yet, in order to use a Kalman filter, additional processing of the data is necessary. Kalman filters only represent the state of the system using a single Gaussian and our data may contain multimodal distributions from

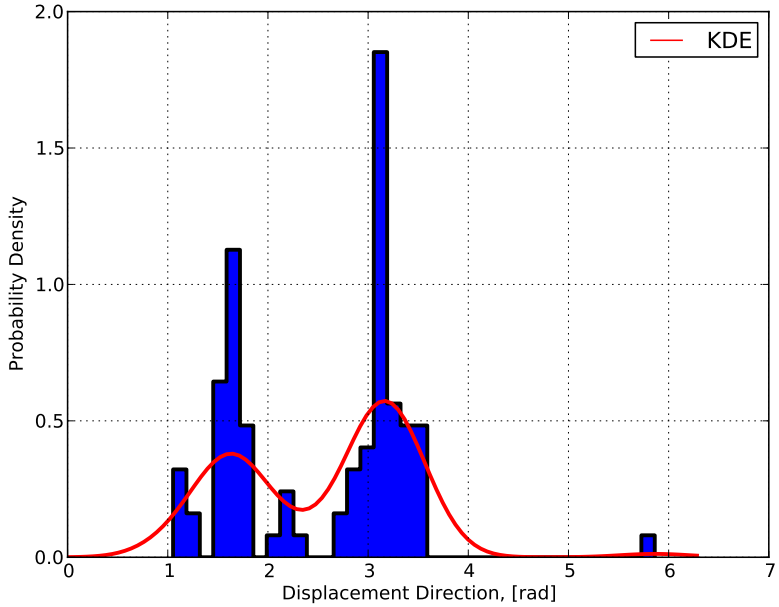


Figure 2. Histogram of optical flow measurements for estimated direction of travel. A Gaussian kernel density estimation is depicted. The histogram indicates a multimodal distribution. The results around 1.5 radians with the lower probability density is the correct result.

erroneous data. Our simplistic solution is to eliminate the multimodal nature of the data by selecting only the highest probability Gaussian representation contained within the data set prior to incorporation into the Kalman filter. Also, only the mean of the resulting measurements contained within a single image is used by the Kalman filter for estimating the direction and magnitude of motion associated with the optical flow measurements.

For the case depicted in Figure 2, the most likely answer from strictly a probability density perspective yields the incorrect result. Kalman filters can effectively smooth through occasional erroneous data values. Yet it can be extremely difficult to recover from the mistake. Particle filters are likely to be more suitable for this application than the traditional Kalman filter approach. Particle filters allow for multiple hypotheses to be tracked and therefore better cope with measurements consisting of multiple modes. Therefore, if it is desired to have a system with more robust properties, a particle filter approach should be considered.

3.5 Localization

The ultimate goal of this work is to use visual odometry measurements in conjunction with low-cost MEMS based inertial sensors to provide improved localization or navigation information. The focus here

is to first provide localization information using only the sensor input from the visual odometry measurements. Thus, we first use only visual odometry processing to determine the path traveled. A standard linear Kalman filter with position and velocity as states and two measurements is implemented using the routines available in the OpenCV library. For use in the Kalman localization routine, the optical flow measurements are converted to an estimated velocity, which is calculated by dividing by the elapsed time between the acquired frames.

The process for testing the experimental setup includes displaying video results to the user as well as recording the video for later post processing. The position and velocity results of the Kalman filter are depicted in near real-time on the laptop display. The additional processes of displaying the in-time measurements does however add to the computational load. For an actual deployed system, only the computed odometry measurements would need to be made available. Therefore, the processor load and visual odometry update rates are not representative of a deployed system.

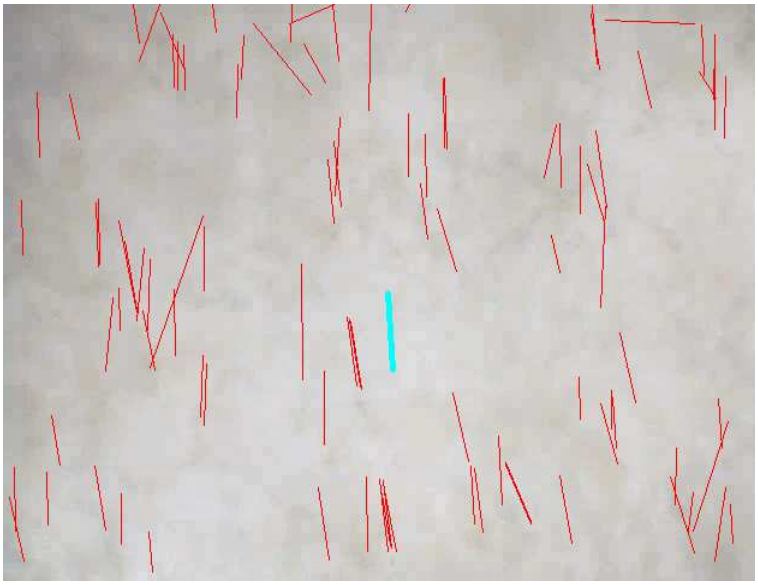
4 Results

The visual odometry experimental setup is used to produce localization results in an indoor environment with a variety of different ground surface textures. The surfaces include both linoleum and short pile carpet surfaces, both with limited indoor lighting conditions. The visual odometry system is capable of sensing a translation within the image frame and is not used to sense rotations. Since an inertial unit is not currently used to sense rotational motion, the laboratory cart motion consists only of simple two-dimensional translations. The majority of the data runs consist of moving the cart in nearly a straight line for a pre-determined distance. Most data runs are for approximately 4.5 m, which is limited by the room size.

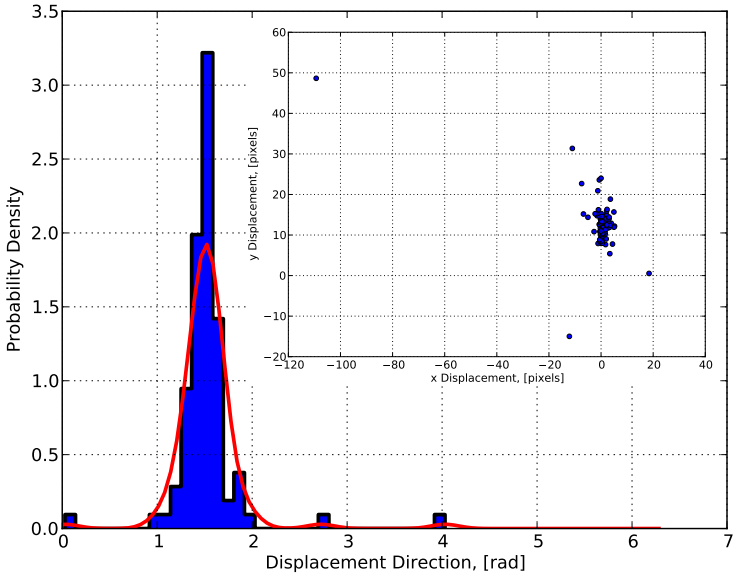
Using the web camera to acquire image frames, feature detection and optical flow results are produced. Figure 3 shows a representative image with unfiltered optical flow measurements and the corresponding sample histogram. The optical flow results are then incorporated into the Kalman filter for position estimation. Figure 4 depicts the estimated localization information using only the web camera for visual odometry measurements.

5 Discussion

The results depicted in Figure 4 indicate that visual odometry measurements from a single downward-pointing camera are useful for determining localization information. Careful inspection of the localization results indicate the measurements as is are not currently robust enough



(a) Optical flow results on a textured surface. The red lines depict offsets for tracked features. The bold cyan line in the middle indicates the estimated result of all the tracked features in the image.



(b) Histogram of optical flow measurements for estimated direction of travel. An x-y scatter plot of the optical flow displacement measurements is shown within the subplot.

Figure 3. Optical flow results.

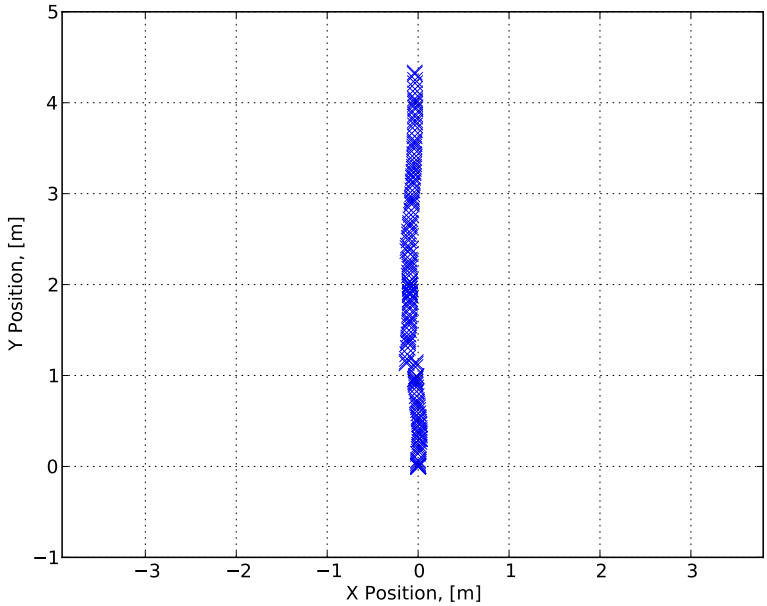


Figure 4. Localization results from visual odometry.

to provide stand-alone high-precision localization information. Yet the results are accurate to a modest level and the results represent the general path traveled during the experiment. Figure 4 does show an irregular discontinuity in the localization information near the x-y position of (0,1) m. For this experiment, the cart input translation is not tightly controlled as the setup is pushed by a human. Sudden changes in input displacement and speed are expected and may be correlated with the human stride. The use of caster wheels on the laboratory cart further allows for cross track translations. Still, a discontinuity in the localization measurements should not be expected. The sudden change in the position results is likely due to an incorrect optical flow measurement getting passed into the Kalman estimation algorithm. Further work to improve the estimation technique is likely to reduce errors due to erroneous measurements.

5.1 Challenges

Although the optical flow measurements are useful for localization information, a few challenges still exist for a potential designer. For instance, valid optical flow measurements are not available at a constant rate from the hardware. The loss of the optical flow measurements will occur. In this experiment, loss of optical flow measurements are often a result of lack of features to track or over/under exposed image frames. The web camera used in this setup adaptively changes the exposure set-

tings and occasionally produces over exposed images and the low-light performance of the hardware is not exceptional. Control over the hardware settings, such as focus and exposure settings will help to control such variations. Still, it should be expected that due to the nature of the imaged surface, adequate tracking features may not exist in the image frames. Thus, using the visual odometry measurements to estimate position by simply summing up the displacement measurements in the along-track and cross-track directions will be in error. The technique applied here, where the displacement measurements are converted to an instantaneous velocity measurement and then used within an estimation algorithm does allow for the occasional missing measurement.

As with any visual processing algorithms, the computational intensity of the routines is always a concern. Processor load is a function of the number of tracked objects, image frame rate and image size. For faster dynamics, higher frame rates are needed. For instance, if the system is only capable of a one frame per second acquisition rate, then the total field of view associated with the camera must not have been traveled during that one second time frame. Depending on the field of view of the camera, this could be a matter of centimeters. By moving the camera further away from the target, the rate of displacement may be increased, but it comes at the cost of reduced resolution and hence fewer tracked objects. With the in-time position visualization disabled, update rates in excess of 10 Hz are achieved with the web camera setup.

For a proof of concept experiment, the utilized hardware is sufficient. Yet, the resulting images are often blurred, limiting the motion to unacceptably slow speeds. The HD camera exhibited much better performance. The resulting video contained more features, exhibited better low-light performance and the resulting images were rarely blurred. The quality of the optical flow results, and hence the derived navigation position information, are substantially better for the HD video case. Unfortunately, the HD video camera could not be tested in near real-time as the available on-hand hardware only records the captured video to a local storage medium.

6 Summary

This work demonstrates the use of a single downward-pointing camera and visual odometry techniques for localization. The technique uses feature detection and optical flow measurements to provide sensor information to localization algorithms. The application is specifically targeted to robotic platforms in GPS-denied environments. The work is primarily intended to provide a proof-of-concept demonstration of the technique and shows potential to aid localization algorithms. Future work will investigate the inclusion of the visual odometry measurements with MEMS based inertial sensors.

References

1. Cheng, Y.; Maimone, M.; and Matthies, L.: Visual Odometry on the Mars Exploration Rovers. *IEEE Conference on Systems, Man and Cybernetics*, IEEE, 2005.

2. Maimone, M.; Cheng, Y.; and Matthies, L.: Two years of visual odometry on the Mars Exploration Rovers. *Journal of Field Robotics, Special Issue on Space Robotics*, vol. 24, 2007.

3. Zaman, M.: High Precision Relative Localization Using a Single Camera. *Proceedings 2007 IEEE International Conference on Robotics and Automation*, IEEE, April 2007, pp. 3908–3914.

4. Shi, J.; and Tomasi, C.: Good Features to Track. *Proceedings Computer Vision and Pattern Recognition, 1994 IEEE Computer Society Conference*, IEEE, 1994, pp. 593–600.

5. Bouguet, J.-Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker. Description of the Algorithm. , Intel Corporation, 2000.

6. Lucas, B. D.; and Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. *Proceedings 7th International Joint Conference on Artificial Intelligence (IJCAI)*, 1981, pp. 674–679.

7. Fischler, M. A.; and Bolles, R. C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, vol. 24, no. 6, 1981, pp. 381–395.

8. Torr, P.; and Zisserman, A.: MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding*, vol. 78, no. 1, 2000, pp. 138–156.