# Analysis of the CMU Localization Algorithm Under Varied Conditions

Aayush Bansal, Hernán Badino, Daniel Huber

CMU-RI-TR-15-05

Jan 2015

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213



This research was sponsored by the Volkswagen Group of America Electronics Research Lab.

# Abstract

Localization is a central problem for intelligent vehicles. Visual localization can supplement or replace GPS-based localization approaches in situations where GPS is unavailable or inaccurate. Although visual localization has been demonstrated in a variety of algorithms and systems, the problem of how to best configure such a system remains largely an open question. Design choices, such as "where should the camera be placed?" and "how should it be oriented?" can have substantial effect on the cost and robustness of a fielded intelligent vehicle. We have previously developed a visual localization algorithm that has been analyzed with respect to seasonal variations and certain environmental conditions. This report extends this analysis with greater detail on how different sensor configuration parameters and environmental conditions affect visual localization performance with the goal of understanding what causes certain configurations to perform better than others. We ground the investigation using extensive field testing of our visual localization algorithm.

# Contents

# 1 Introduction

Localization is a central problem in robotics – a problem that must be addressed in many intelligent vehicle applications. For example, visual SLAM algorithms (e.g., [3, 4, 7]) use localization to dynamically build up a map and detect loop closures, while commercial automotive navigation systems and visual localization algorithms (e.g., CMU's GPS-denied visual localization algorithm [2]) determine a vehicle's position with respect to a prior map.

While localization can often be accomplished using GPS, an autonomous vehicle must be robust to situations where GPS is unavailable, such as when driving downtown in large cities or along forested, rural roads. Furthermore, localized GPS jamming is becoming increasingly common. As a consequence, fielded autonomous vehicles will likely need to incorporate alternative sensing modalities for localization. Appearance-based (i.e., visual) localization is one promising candidate.

We have previously developed a GPS-denied visual localization algorithm and analyzed it with respect to seasonal variations and limited environmental conditions. In order to determine the feasibility of deploying the algorithm in commercial systems, a more detailed analysis is needed. The goal of this work is to gain an understanding of how different sensor configuration parameters and environmental conditions affect this algorithm. We conducted experiments using the Navlab experimental testbed vehicle, which we outfitted with a Ladybug 5 omnidirectional camera. We defined a test route consisting of sequences of different environment types, and we collected data under varying environmental conditions and with different system configurations. We then evaluated the performance of our algorithm in terms of accuracy, failure rate, and computational requirements. The results of our analysis indicate that the algorithm reliably localizes the vehicle across a variety of environmental conditions and system configurations, with an accuracy of 1 to 2 meters on average. The best configuration aims the camera at an angle of about 67.5° with respect to the front of the vehicle (or alternatively, masks out regions in the front of the vehicle if the camera is pointing forward). We also found that the algorithm degrades gracefully with lower map resolutions and that it performs well even with image resolutions as low as 24x24 pixels.

# 2 Hardware Setup

Our data collection testbed is based on the Navlab 11 autonomous vehicle (Figure 1a). Among other things, the vehicle is equipped with an IMU, GPS, and a computing infrastructure that enables real-time data synchronization and logging. We augmented

the baseline platform with a Point Grey Ladybug 5 panoramic camera mounted above the hood (Figure 1b). The camera captures six 2448 x 2048 images at 10 Hz. The images can be stitched together into a panorama that covers 90% of the viewing sphere.

The benefit of the panoramic camera is that it enables the comparison of different viewing directions and fields of view for the exact same data sequence. Virtual video streams are extracted from the panoramic video by cropping different sections of the stitched spherical panorama and rectifying the image following a pinhole model. One downside to this virtual camera approach is that the position of the virtual camera is limited to the position of the actual camera, so, for example, it is not possible to generate an image for a virtual camera mounted on the front bumper.

After data collection was complete, we discovered a problem with the synchronization between the computer controlling the Ladybug 5 camera and the computer controlling the GPS/IMU. As a result, the synchronization between imagery and vehicle ground truth position was not correct. We corrected the problem by resynchronizing the data using manually determined correspondences between the two data streams. Specifically, we identified points where the vehicle came to a stop or began moving, both of which can be readily identified in each data stream, and then applied a linear correction to the timestamps on the imagery.

The synchronization problem was later diagnosed as a deficiency in the design of the Network Time Protocol (NTP) client on the Windows operating system. The problem has since been corrected, but it was not practical to retake all the data for this analysis. The result of the problem is that our analysis is likely overestimating the error in positioning. We do not feel that this is a significant problem, since the localization accuracy levels we achieved, even with this additional error, were good. Furthermore, there are other sources of error of the same magnitude, which we cannot control (e.g., GPS drift).

# 3 Experimental Plan

We identified four different configuration parameters and two environmental conditions that we considered likely to affect the algorithm's performance. For each parameter or condition, we identified a set of values to be tested. The full list of parameters and conditions, along with the values tested, is shown in Table 1.

Based on the target conditions and parameters, we established a test route in the vicinity of CMU that would fully exercise these conditions. In particular, we devised the route so that it contained extended contiguous regions of a given environment type. The chosen route consists of an initial rural/park-like section, followed by a

(a) (b)

Figure 1: (a) The Navlab 11 data collection testbed used in this study. (b) A close-up of the mounting of the Ladybug 5 panoramic camera.

Figure 2: An example panoramic image obtained by the Ladybug 5.

| **Parameter** | **Values** |
|---|---|
| Camera orientation | 0°, 22.5°, 45°, 67.5°, 90° |
| Field of view | 45°, 60°, 75°, 90° |
| Image resolution | 24x24, 48x48, 96x96, 192x192, 384x384 pixels |
| Map resolution | 0.25, 0.5, 1, 2, 3, 6, 10, 20, 30 meters |
| Environment setting | rural, commercial, residential |
| Time of day | noon, afternoon, evening, partly cloudy, cloudy |

Table 1: Configuration and environmental parameters and their values

| Dataset | Morning | Partly Cloudy | Afternoon | Evening | Noon | Cloudy Day |
|---|---|---|---|---|---|---|
| Route length (Km) | 16.2 | 16.4 | 16.3 | 16.3 | 12.3 | 12.4 |
| Date (in 2013) | Oct 8 | Oct 8 | Oct 8 | Oct 8 | Oct 9 | Nov 7 |
| Time of day | 9:54 – 10:28 | 12:44 – 13:21 | 15:21 – 16:04 | 18:08 – 18:47 | 12:07 – 12:39 | 14:46 – 15:14 |

Table 2: Dataset description

commercial section, and concluded by a residential section. The total route length is 16 km and takes approximately 35 minutes to drive. The route map is shown in Figure 3.

We traversed the test route multiple times in the same day to obtain the time of day variations, and on separate days, as weather permitted, to obtain the sunlight variations. Table 2 summarizes the data collection trials used for this analysis.

# 4 Analysis of Configuration Parameters

We divide our analysis into factors that can be controlled by the developer and those that are outside of the developers control. In this section, we analyze the design choices related to sensor configuration: region selection, orientation, field of view, image resolution, and image rate. In each experiment, all of the parameters/conditions are set to their baseline values (as shown in Table 1), and only the parameter/condition being tested is varied.

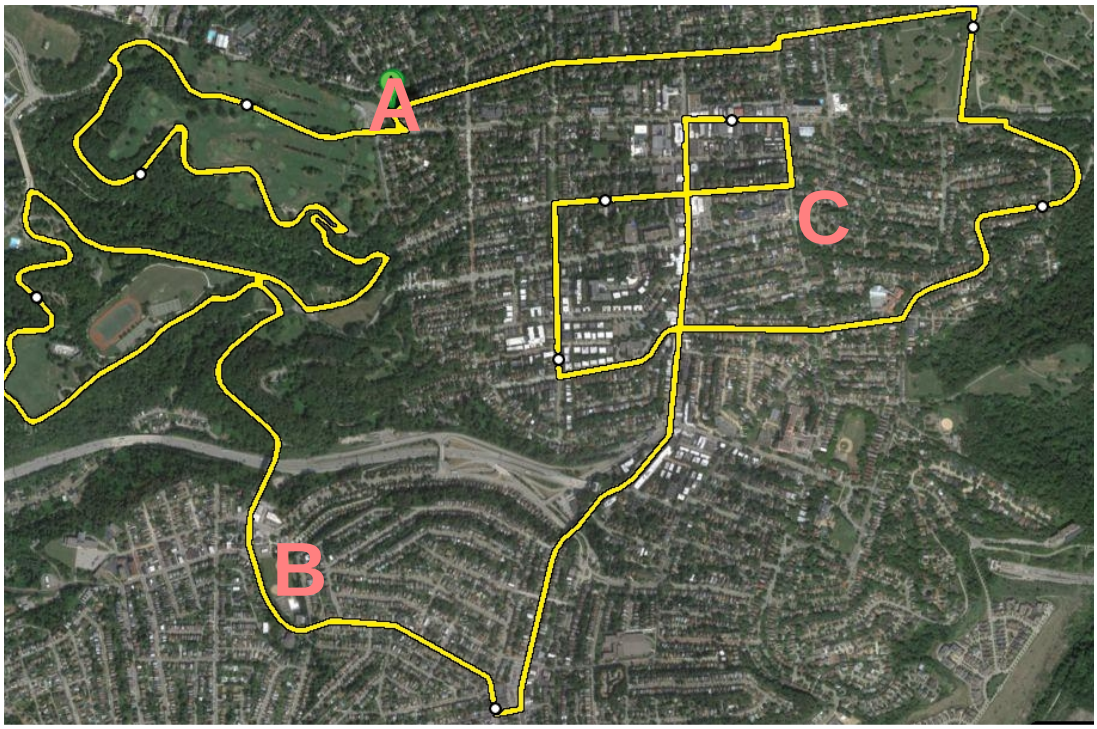



Figure 3: The test route contains sections of rural/park (marker A to B), urban (B to C), and neighborhood (C to A) environments.

## 4.1 Region Selection

We conducted an initial experiment on region selection using previously collected data. In that experiment, we found that using a forward-facing camera, masking out the sky, road and vehicle hood improved localization performance. Based on these results, we reformulated the region selection problem in an information theoretic framework.

Consider an image taken from a forward-looking camera on a vehicle. A typical scene contains regions of road, sky, and traffic ahead, and buildings, trees, parked cars, and signs on the sides. Intuitively, unique landmarks, such as particular buildings, are more informative for determining location than generic objects (e.g., trees, road, and sky) or moving objects (e.g., traffic). Therefore, we hypothesize that a good sensor configuration for localization will focus the sensing on the sides of the road. However, side-facing sensors could actually perform worse due to mitigating factors, such as motion blur and the limited time that objects are within the sensor's field of view.

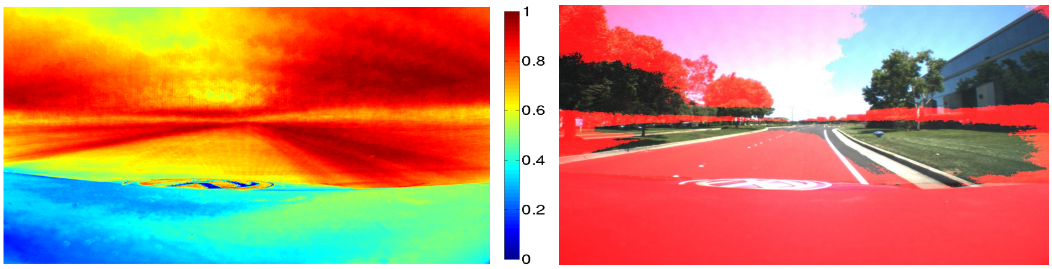
(a) (b) (c) (d) (e) (f)

Figure 4: Examples of different environment types: rural/park (a-b), commercial (c-d), and residential (e-f).

We can use information theory to formalize and mathematically quantify how unique or interesting an image region is. It is well-known from information theory that entropy is a measure of the amount of information in a distribution. Here, we compute the entropy of each pixel location over all the images from an entire route.

$$H = -\sum_i p_i log(p_i),$$

where $p_i$ is the probability distribution of grey-scale intensity values at each pixel location. The resulting entropy-based heat map, normalized to [0,1] in Figure 5, shows




(a) (b)

Figure 5: (a) The heat map showing the most informative regions of the view (b) Overlaying the mask onto the original image and thresholding low values shows that the informative regions lie on the sides of the road.

Figure 6: Region selection using camera orientation = 45° and FOV = 90°.

graphically that the left and right sides of the road have higher entropy, and therefore contain more information than the other regions, such as the hood, sky, and the road itself. Similar patterns were observed in other data sets that used different camera configuration settings.

We tested the performance of region selection using virtual cameras with 45° orientation from front-facing, and 90° field of view (FOV). Using manually selected regions (Figure 6) to discard the sky and road results in a reduction of localization error in every test case (Figure 7 and Table 3). We also analyzed the number of divergences, which we define as instances during which the vehicle's position estimate error exceeds 10 m. Using region selection reduces the number of divergences to zero in every test case.

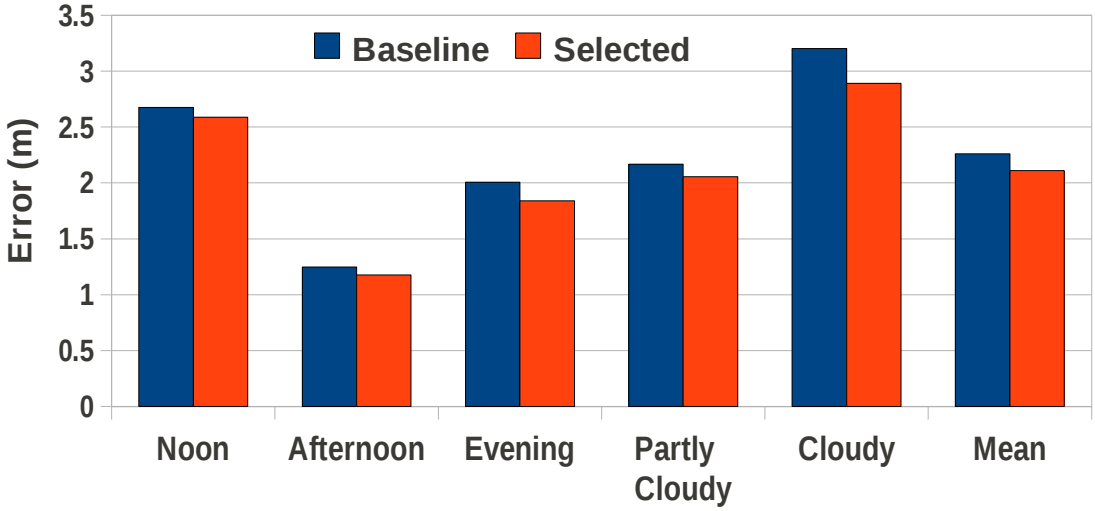



Figure 7: Result of using region selection shown in Figure 6 using the database from the morning trial.

| | **Avg Error (#Div)** | |
|---|---|---|
| **Time of Day** | **Baseline** | **Region Selection** |
| Noon | 2.67 (0) | 2.59 (0) |
| Afternoon | 1.25 (1) | 1.18 (0) |
| Evening | 2.00 (2) | 1.84 (0) |
| Partly Cloudy | 2.16 (2) | 2.06 (0) |
| Cloudy | 3.20 (2) | 2.89 (0) |
| **Mean** | 2.26 (1.4) | 2.11 (0) |

Table 3: Region selection. Eliminating non-informative regions reduces localization error and divergences.

## 4.2 Camera Orientation

In this section, we analyze the effect of camera pointing direction on performance. We used images extracted from the panoramic camera image, varying the virtual camera orientation from 0° (front-facing) to 90° in increments of 22.5° while keeping the FOV constant at 90°. The results, shown in Figure 8 and Table 4, show that accuracy improves steadily with increasing orientation toward the side, but the improvement between 67.5° and 90° is minimal. Looking at the original images, these last two orientations contain minimal amount of low-information regions (i.e., road and sky).

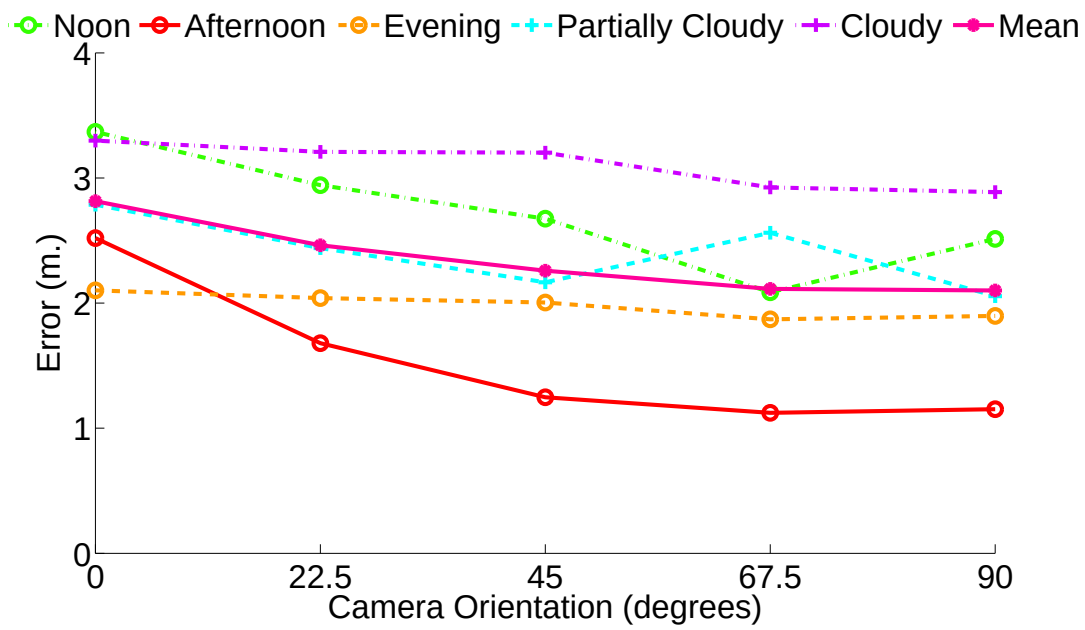



Figure 8: The effect of camera orientation on localization accuracy.

These orientations also result in the fewest divergences.

## 4.3 Field of View

Field of view (FOV) can potentially affect localization performance. Increasing the FOV observes a wider area and potentially more distinctive locations. However, for a given image resolution, a wider FOV reduces the detail at the pixel level. It is not obvious which factor is more influential. An additional question is whether the best FOV choice is the same for different camera orientations.

To test these ideas, we conducted an experiment similar to the one for camera orientation. Using the panoramic camera, we extracted sub-images with FOV values ranging from 45° to 90° in 15° increments. We used the same data sets as in the orientation experiment and, in fact, tested all combinations of orientation with each FOV value. The results, summarized in Table 4, show that the best FOV choice depends somewhat on the camera orientation – a smaller FOV for the 45° orientation performed best, whereas the 90° FOV performed best on the 90° orientation. Looking at the raw images, the wider FOV values at 45° orientation capture significant (non-informative) areas of road and sky, which explains why narrower FOVs perform

| Camera Parameters | | Time of Day | | | | | |
|---|---|---|---|---|---|---|---|
| **Orientation** | **FOV** | **Noon** | **Afternoon** | **Evening** | **Partly Cloudy** | **Cloudy** | **Mean** |
| 0 | 90 | 3.36 (2) | 2.51 (1) | 2.10 (4) | 2.78 (4) | 3.29 (3) | 2.81 (2.8) |
| 22.5 | 90 | 2.94 (0) | 1.67 (1) | 2.03 (2) | 2.43 (2) | 3.20 (1) | 2.46 (1.2) |
| 45 | 45 | 2.58 (0) | 1.22 (0) | 1.85 (0) | 2.06 (0) | 2.89 (0) | 2.12 (0) |
| 45 | 60 | 3.19 (1) | 1.21 (0) | 1.86 (0) | 2.10 (0) | 2.88 (0) | 2.25 (0.2) |
| 45 | 75 | 2.66 (0) | 1.20 (1) | 1.95 (4) | 2.10 (0) | 3.10 (2) | 2.20 (1.6) |
| 45 | 90 | 2.67 (0) | 1.24 (1) | 2.00 (2) | 2.16 (2) | 3.20 (4) | 2.25 (2.8) |
| 67.5 | 45 | 2.57 (0) | 1.18 (0) | 1.89 (0) | 2.08 (0) | 2.90 (0) | 2.12 (0) |
| 67.5 | 60 | 2.55 (0) | 1.18 (0) | 1.88 (0) | 2.08 (0) | 2.86 (0) | 2.11 (0) |
| 67.5 | 75 | 2.55 (0) | 1.14 (0) | 1.87 (0) | 2.07 (0) | 2.89 (0) | **2.10 (0)** |
| 67.5 | 90 | 2.08 (0) | 1.12 (1) | 1.87 (1) | 2.08 (0) | 2.92 (0) | 2.11 (0.4) |
| 90 | 45 | 2.54 (0) | 1.29 (0) | 1.93 (0) | 2.05 (0) | 2.93 (0) | 2.15 (0) |
| 90 | 60 | 2.52 (0) | 1.22 (0) | 1.91 (0) | 2.06 (0) | 2.87 (0) | 2.12 (0) |
| 90 | 75 | 2.52 (0) | 1.19 (0) | 1.91 (0) | 2.05 (0) | 2.86 (0) | 2.11 (0) |
| 90 | 90 | 2.51 (0) | 1.15 (1) | 1.89 (0) | 2.05 (0) | 2.88 (0) | 2.10 (0.2) |

Table 4: Localization error as a function of camera FOV for a given camera orientation

better at this orientation. Aside from this, the effect of different FOV values overall is modest, suggesting that the information content is distributed evenly across scales.

## 4.4 Image Resolution

The choice of image resolution determines the quality of camera (and, therefore, the cost) needed for robust localization. In this experiment, we varied the resolution by factors of two from 384 x 384 pixels down to 24 x 24 pixels. The results, shown in Figure 9 and Table 5, show that reducing the image resolution does not have much of a detrimental effect on localization. Even down to a resolution of 24 x 24 pixels, there is almost a negligible gain in localization error. This suggests that even a very basic camera could be used for visual localization and indicates that visual localization could scale to extremely large data sets very effectively. This result is not too surprising, since images at this resolution have also proven useful for recognition and classification tasks [9]. Note that the good performance at low resolution may be tied to the image descriptor used for the localization.

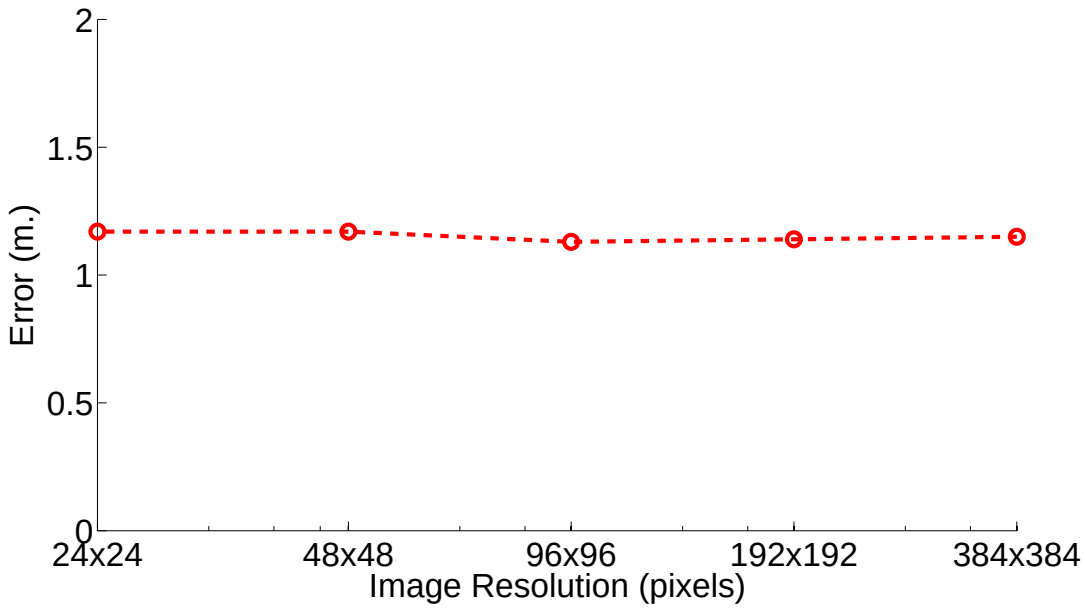



Figure 9: Effect of image resolution

| | Error (m) | | | |
|---|---|---|---|---|
| **Image Resolution** | **Avg** | **Max** | **Stddev** | **#Div** |
| 384x384 | 1.15 | 0.93 | 5.52 | 0 |
| 192x192 | 1.14 | 0.93 | 5.48 | 0 |
| 96x96 | 1.13 | 0.92 | 5.45 | 0 |
| 48x48 | 1.17 | 0.98 | 6.88 | 0 |
| 24x24 | 1.17 | 0.99 | 6.50 | 0 |

Table 5: Effect of image resolution on localization accuracy. Camera Orientation = 67.5° and FOV = 75°.

## 4.5 Image Frequency / Map Resolution

Image frequency affects the cost of a localization system in terms of the quality of camera needed (i.e., high frame-rate vs. low frame-rate) and the amount of storage needed for the map representation, both of which affect the cost of a fielded system. To quantify the effect of image frequency, we vary the map resolution, i.e., the physical distance between nodes in the map. While map resolution and frame rate are not equivalent, the effect of lower frame rates would be expected to be similar to that of a lower map resolution.

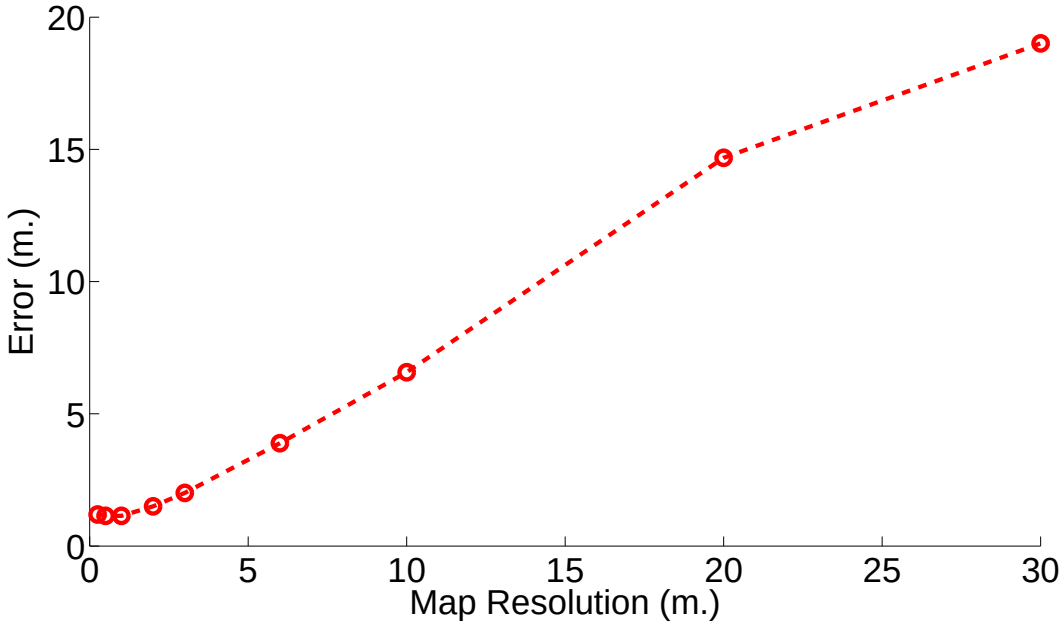



Figure 10: Effect of map resolution on localization accuracy.

To analyze the effect of map resolution, we varied the distance between map nodes between 0.25 m to 30 m. The results, shown in Figure 10 and Table 6, show that the accuracy increases approximately linearly with increasing map resolution up to 1 m resolution, at which point no more improvement is seen. This limit on the maximum accuracy is probably related to the accuracy of the ground truth in these data sets, which is limited by the accuracy of the GPS/IMU used by the testbed vehicle. While the exact localization accuracy will be algorithm-dependent, the relationship between accuracy and map resolution suggests an upper limit on the required camera frame rate. For example, if the maximum vehicle speed is 80 MPH ($\sim$35 m/s), a frame rate of 7 fps will ensure one image is captured every 5 m and the expected average localization accuracy would be 2.5 m.

# 5 Environmental Factors

The previous section focused on configuration parameters and how they affect localization accuracy. In this section, we address factors that are beyond the control of the system designer. While there may be nothing that can be done from a system

| | Error (m) | | | |
|---|---|---|---|---|
| **Map Resolution** | **Avg** | **Max** | **Stddev** | **#Div** |
| 0.25 | 1.19 | 4.30 | 0.87 | 1 |
| 0.5 | 1.14 | 4.44 | 0.86 | 0 |
| 1 | 1.14 | 5.48 | 0.93 | 0 |
| 2 | 1.50 | 16.62 | 1.55 | 0 |
| 3 | 2.01 | 28.03 | 2.66 | 0 |
| 6 | 3.89 | 52.85 | 5.27 | 0 |
| 10 | 6.57 | 81.00 | 8.94 | 0 |
| 20 | 14.68 | 99.93 | 16.90 | 4 |
| 30 | 19.01 | 99.90 | 18.76 | 7 |

Table 6: Effect of map resolution on localization accuracy. Camera Orientation = 67.5° and FOV = 75°.

standpoint with respect to these factors, it is nevertheless useful to understand how they affect localization performance because these insights can drive development of new and potentially more robust localization methods. Specifically, we consider three environmental factors: environment type, time of day, and sunlight conditions. The effect of seasonal variation has been studied previously [2, 10].

## 5.1 Environment Type

Different environmental settings offer different amounts and quality of visual features. For example, we would expect rich visual features in commercial or downtown areas but sparse features in rural areas. It is an open question whether localization is easier or harder in specific types of environments. An understanding of what environment types are most challenging can provide insight into where a localization algorithm is most likely to break down.

To analyze the effect of environment type on localization accuracy, we separated the data from the test route into three basic environmental types: rural/parklike, commercial/downtown, and residential/suburban. In fact, the test route was chosen to provide distinct, contiguous segments according to these categories. The results, shown in Table 7, are somewhat surprising. While the rural/parklike segment proved to be the least accurate, the residential/suburban segment accuracy was actually better than the commercial segment. One possible reason for this unexpected result is that the commercial areas have significantly more traffic and also more variability on the roadside due to changing parked cars in different trials. Another explanation may be that

| | Error (m) | | | |
|---|---|---|---|---|
| **Environment** | **Avg** | **Max** | **Stddev** | **Div** |
| Rural | 1.45 | 5.26 | 1.04 | 0 |
| Commercial | 1.14 | 4.09 | 0.80 | 0 |
| Residential | 1.07 | 4.68 | 0.82 | 0 |

Table 7: Effect of environment type on localization accuracy.

the commercial areas, while more structured and feature rich also have more repetitive patterns, such as repeating windows on buildings, which can cause increased uncertainty at a fine-grain scale when localizing.

## 5.2 Time of Day and Sunlight Conditions

The time of day can have a significant effect on the appearance of a scene, particularly on sunny days. The appearance of shadows, which change over time, can dominate more stable features in the environment. The sun position during early mornings and late afternoons or evenings can create artifacts in the image due to sensor saturation, and the extreme brightness variations can exceed the dynamic range of standard cameras. Furthermore, factors not related to illumination can affect the appearance of the environment. For example, typically, commercial areas will experience less traffic in the mid-morning than in the afternoon, when localization may be more challenging due to additional clutter from parked cars, people, and traffic.

We studied the effects of time of day and lighting conditions by conducting trials at various times of the day on a sunny day (morning, noon, afternoon, and evening) and on cloudy and partly cloudy days. We used the morning trial for creating the map and compared localization accuracy across the remaining trials. The results of the experiment, shown in Figure 8 show, for the most part, what intuition suggests. The bigger the difference in time of day, the more challenging the localization task becomes. The graph includes two surprises, however. First, the localization error for the noon trial was slightly higher than the afternoon one. This could be due to the fact that the noon trial was conducted on the next day due to changing weather conditions. The second surprise was that cloudy and partly cloudy conditions were more challenging to match with the sunny morning map than different times of day under sunny conditions. Intuitively, no shadows would match morning shadows better than afternoon shadows.

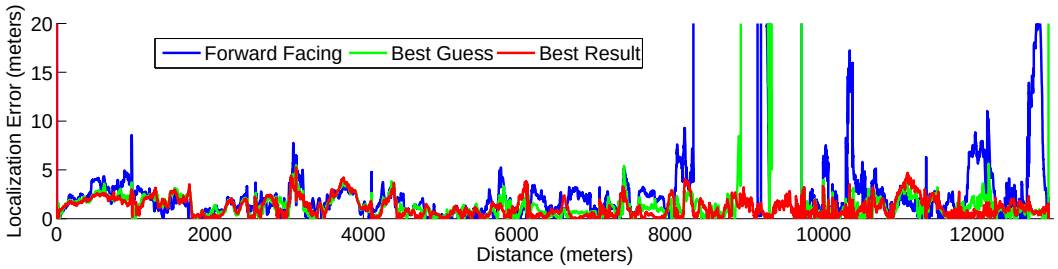



Figure 11: Comparison of the naive, front-facing camera (forward facing), initial best guess configuration (best guess) with the best-performing configuration determined from the analysis (best result).

# 6 Summary and Conclusions

We have studied how the choice of sensor configuration parameters and how various environmental factors affect the performance of visual localization. We conducted an extensive series of experiments using both forward-facing cameras and virtual cameras extracted from panoramic imagery. Using an information-theoretic approach, we established a relationship between the information content of image regions and the usefulness of those regions for localization. Our findings reflect the intuition that the sides of the road provide the most benefit for localization algorithms. Interestingly, many visual localization and mapping algorithms focus primarily on forward-looking cameras [1, 5, 6, 8]. Our findings suggest that a better approach would be to point the cameras at an angle of 45° or more away from front-facing, or at least masking out the road and sky regions of the image. We compared our initial choice for the best performance of the localization algorithm used in this paper with the revised choice based on the insights from this analysis (orientation = 90°, FOV = 75°, image resolution = 192x192, map resolution = 1 m). The results (Figure 11) show significant improvement, with a reduction in average localization error of 4% and elimination of all temporary divergences, and a reduction over the error of a front-facing camera of 53%.

# References

[1] S. Achar, C. V. Jawahar, and K. M. Krishna, "Large scale visual localization in urban environments," in *ICRA*, 2011.

[2] H. Badino, D. F. Huber, and T. Kanade, "Real-time topometric localization," in *ICRA*, 2012, pp. 1635–1642.

[3] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, 2008.

[4] M. Cummins and P. M. Newman, "Appearance-only slam at large scale with fab-map 2.0," *I. J. Robotic Res.*, vol. 30, no. 9, 2011.

[5] A. J. Glover, W. P. Maddern, M. Milford, and G. F. Wyeth, "FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day," in *ICRA*, 2010.

[6] A. H. A. Hafez, M. Singh, K. M. Krishna, and C. V. Jawahar, "Visual localization in highly crowded urban environments," in *IROS*, 2013.

[7] M. Milford and G. F. Wyeth, "Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights," in *ICRA*, 2012.

[8] P. Neubert, N. Sünderhauf, and P. Protzel, "Appearance change prediction for long-term navigation across seasons," in *ECMR*, 2013.

[9] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, 2008.

[10] C. Valgren and A. J. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, no. 2, pp. 149–156, 2010.