Understanding How Camera Configuration and Environmental Conditions Affect Appearance-based Localization

Aayush Bansal, Hernán Badino, Daniel Huber
The Robotics Institute, Carnegie Mellon University
{aayushb,hbadino,dhuber}@cs.cmu.edu

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. This paper analyzes 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 and providing general principles for configuring systems for visual localization. We ground the investigation using extensive field testing of a visual localization algorithm, and the data sets used for the analysis are made available for comparative evaluation.**

I. 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., [7], [8], [17]) use localization to dynamically build up a map and detect loop closures, while commercial automotive navigation systems and visual localization algorithms (e.g., topometric localization [6]) 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.

Although visual localization has been demonstrated in a variety of algorithms and systems [1], [5], [6], [12], [19] the problem of how to best configure such a system remains largely an open question. When developing a visual localization algorithm, a designer is faced with a number of trade-offs in terms of choice of camera, number of cameras, configuration of optics, sensor orientation, and frame rate. Will the algorithm be robust with a lower resolution, hence lower cost, camera? How wide should the camera's field of view be? How many cameras are necessary, and which way should they be oriented? Such choices can have a substantial effect on the cost and robustness of a fielded product.

The goal of this paper is to gain an understanding of how different sensor configuration parameters and environmental conditions affect visual localization algorithms. By studying these effects, we can not only better understand how to configure a visual localization system, but also gain understanding of the underlying reasons that some configurations perform better than others. We will ground our investigation using extensive field testing of a visual topometric localization algorithm [5]. The conclusions, however, could be generalized to other algorithms that employ visual localization [8], [14]. The key contributions of this paper are: 1) a detailed study of the relationship between system configuration and visual localization performance; 2) an analysis of the underlying causes for certain regions of the viewing sphere around a vehicle to be more effective for localization; and 3) a large-scale, publicly available data set of panoramic imagery with ground truth localization for the research community.

II. Related Work

Localizing an observer by means of evaluating the acquired images of the environment is a technique that has been widely explored in the literature. The most used techniques, such as SLAM [9], [13] and visual odometry [4], rely on geometric constraints on the structure of the world and its projection onto the images. These methods usually perform feature tracking, triangulation, and pose estimation, providing metric information of the location of the cameras. Alternative approaches to localization deal with place recognition [8], [23], where localization is achieved by identifying previously visited locations. These methods extract features from the images and create a topological map of the environment modeled by a graph. Features from the current view are then matched to those in the map to identify the location of the observer within the graph.

Despite the large amount of previous work on visual localization, relatively little of this work explores the robustness of the algorithms to various changes in the environment. Typical test data is gathered all at one time [7] or over a period of a few days [8] without controlling for such factors as illumination. However, some researchers have explored the effects of time and environmental variation on localization and place recognition. Glover et al. [11] measured the performance of various SLAM algorithms using data sets from different times of the day. Valgren and Lilienthal [24] evaluated the performance of SIFT and SURF in the context of long-term
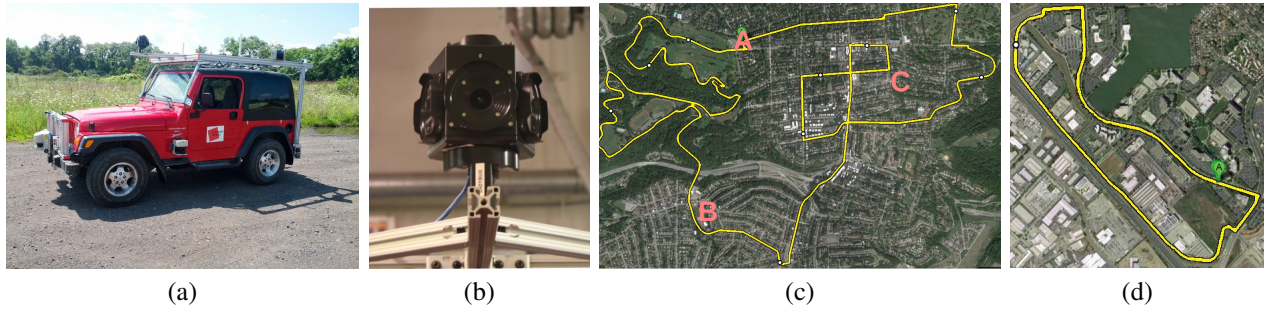
Fig. 1. (a) The Navlab 11 data collection testbed used in this study. (b) A close-up of the mounting of the Ladybug 5 panoramic camera. (c) Test route 1 contains sections of rural/park (marker A to B), urban (B to C), and neighborhood (C to A) environments. (d) Test route 2.

topological localization. Data sets were acquired over long periods of time to capture the natural seasonal changes in an outdoor environment. The authors concluded that SURF performs better than SIFT for the purpose of localization in outdoor environments. Badino et al. [6] also analyzed the localization performance across seasonal variations using a topometric approach. Similarly, Milford et al. [17] developed a robust localization algorithm adding temporal filtering to an image template matching technique. The authors show results on data acquired over different seasons and weather conditions. More recently, Neubert et al. [19] showed that location matching can be improved by actively predicting the effect of seasonal variations.

The effects of sensor configuration choices have been studied only to a limited extent. Typically, systems are designed and configured using trial and error or intuition, and the methods for making the decisions are not normally documented. Vision-based odometry or localization algorithms are usually evaluated using either front-facing cameras [1], [11], [16], [19] or cameras pointed to the side [7], [20], [21] and using average focal length lenses. Some researchers have proposed using panoramic [3], [18] or omni-directional imagery [15], which could potentially improve localization by using a much wider field of view (FOV). However, our analysis shows that a wider FOV does not necessarily improve localization accuracy. An alternative to explicitly configuring the camera orientation and FOV is to actively learn what part of an image contains useful features for localization. Hafez et al. [12] showed that weights for discriminative features (ones which remain consistent both spatially and temporally) can be learned by traversing the same route multiple times. The insights from our work could be leveraged to improve such an approach by identifying the most useful direction to aim the camera's limited viewing area, thus enabling the algorithm to learn better features in fewer iterations.

While visual localization has been studied extensively, there are no standard, publicly available data sets that can be used to compare the effectiveness of different approaches across different environmental and sensor configuration conditions. Several data sets for visual localization have been made available (e.g., KITTI [10] and FAB-MAP [8]), but these data sets were designed for different problems – visual

odometry for KITTI and loop-closure detection for FAB-MAP. To test performance under different environmental conditions, it is necessary to traverse the same route multiple times under different conditions, as is provided by our data set. Furthermore, repeated route traversal is more suitable for testing localization with respect to a prior map.

III. FRAMEWORK

The questions posed in this paper can only be answered using an extensive data set that covers the range of configurations that could be utilized for a localization algorithm as well as the diverse environmental conditions that such an algorithm may face. Ideally, it would be possible to control every parameter and condition independently, but that is challenging, except with synthetic data. Knowing that some aspects of the localization problem would not be modeled with synthetic data, we chose to focus on real-world data collection, with the goal of minimizing differences between data sets with respect to parameters or conditions not being controlled.

A. Data Collection Testbed

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 fps. 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.

Dataset	Morning	Partly Cloudy	Afternoon	Evening	Noon	Cloudy	Bay 1	Bay 2
Route length (Km)	16.2	16.4	16.3	16.3	12.3	12.4	4.4	4.4
Date (in 2013)	Oct 8	Oct 8	Oct 8	Oct 8	Oct 9	Nov 7	July 18	July 19
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	14:00 – 14:11	9:54 – 10:03

B. Data Sets

We first identify the main system configuration parameters and environmental conditions that could affect the performance of a localization algorithm. In this paper, we consider four configuration parameters: camera orientation, camera field of view, image resolution, and video frame rate. We also analyze three environmental parameters: time of day (morning, midday, afternoon, dusk), sunlight (sunny, cloudy), and environment (commercial, residential, rural). We do not consider seasonal variations and adverse weather conditions such as fog, rain, and snow, or nighttime driving, which have been studied previously [6], [17], [24].

We established a consistent data collection route to enable comparison across these different conditions. The route, shown in Figure 1c, is 16 km long and takes approximately 35 minutes to navigate (see Table I). The route consists of distinct environmental segments: rural (a park), commercial (medium-density commercial district), and residential (suburban neighborhoods). We traversed the 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.

The data sets are available online at http://3dvis.ri.cmu.edu to enable researchers to evaluate and compare localization algorithms on standard data sets. As mentioned in Section II, no existing large scale data set for visual localization is available. The data sets each consist of a sequence of time stamped images along with synchronized vehicle location ground truth obtained from the onboard IMU and GPS.

We also evaluate the performance of a purely front-facing camera using two data sets from a dashboard-mounted camera. The two data sets consist of a 4.4 km route (Figure 1d) and were obtained at different times of the day on two consecutive days.

IV. Visual Topometric Localization Overview

In order to ground our analysis of visual localization configuration, we use an existing algorithm known as visual topometric localization in our experiments. For completeness, we briefly review the algorithm and its salient characteristics. Details of the algorithm are found in [5] and [6]. Metric localization, such as those based on SLAM [2] and visual odometry [4] are accurate for short distances, but they drift over time or fail at long sequences, especially if the vehicle does not revisit the same place on its trajectory. On the other hand, topological localization methods, such as [23] estimate the observer's location qualitatively from a finite set of possible positions avoiding localization drift but provide

only rough position estimates. Topometric localization was first introduced in [5] and improved in [6] as a combination of topological and metric localization. In topometric localization, the space of possible locations is discretized and modeled by a graph. Nodes in the graph represent possible locations and edges of the graph connect adjacent locations. Localization is achieved by finding the node associated to the position of the vehicle. Thus, topometric localization provides metric information of the location of the vehicle while avoiding the typical drift of dead-reckoning algorithms.

Topometric localization involves two tasks: creating a map of the environment and localizing with respect to the map. In the map creation stage, a vehicle equipped with GPS and a camera travels the route to be recognized. The location of the vehicle is acquired at regular intervals and a graph of the route is created. At the same time, features are extracted from the acquired images and linked to the created nodes. The map creation can be performed online, since the computation and storage requirements are minimal.

In the localization stage, a vehicle without GPS drives over the mapped routes acquiring imagery data. A Bayes filter is used to estimate the probability distribution of the position of the vehicle by matching features extracted from the images with those already stored in the map database. The Bayes filter further propagates the probabilities within the graph as the vehicle moves between connected nodes. The estimated location of the vehicle is obtained after each measurement from the maximum a posteriori (MAP) probability.

Topometric localization has been shown to be robust to extreme changes in the environment such as those produced by seasonal changes, lighting differences, and occlusions thanks to the appropriate design of the Bayes filter [6]. Furthermore, the computational and storage requirements are very low, making it suitable for very large scale urban localization.

A. Evaluation Criteria

The localization accuracy is computed with respect to the GPS-based ground truth in terms of longitudinal error (i.e., along the route). In this paper, we do not consider lateral error (i.e., across the lanes), since that is not the primary goal of route localization. The ground-truth position and the estimated position are, therefore, projected to the trajectory map to compute the position error along the route.

For each experiment, we report four statistics: average localization error (Avg), maximum localization error (Max), standard deviation (Stddev), and number of divergences (#Div). A divergence is defined as occurring when the localization error exceeds 30 m. In the map resolution
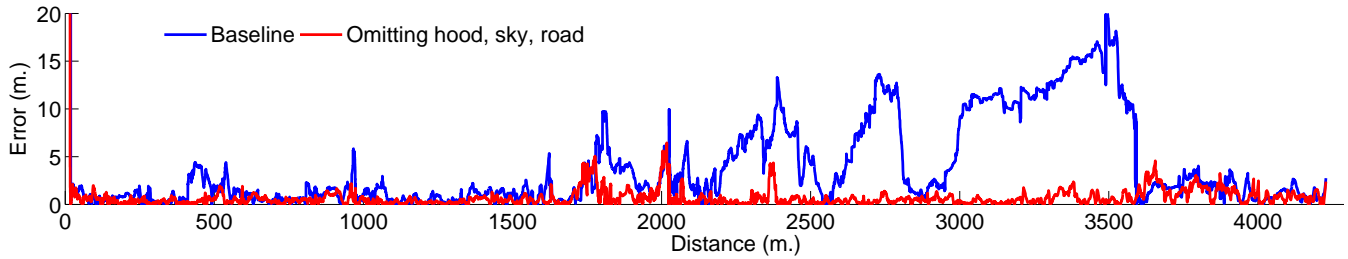
Fig. 2.   Localization accuracy for region selection vs. baseline over an entire route.

experiment (Section V-F), this threshold is set to 100 m to allow testing of image frequencies on the order of 30 m. The remaining three statistics are computed using only non-diverging regions. All errors are reported in units of meters.

V. SENSOR CONFIGURATION ANALYSIS

We divide our analysis into factors that can be controlled by the developer and those that are outside of the developer's control. In this section, we analyze the design choices related to sensor configuration: region selection, orientation, field of view, image resolution, and image rate.

A. An Information-theoretic Viewpoint

Consider an image taken from a forward-looking camera on a vehicle. A typical scene contains regions of relatively unchanging objects, such as road, sky, and the traffic, and more variable objects, such as buildings, trees, 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). Furthermore, objects on the sides of the vehicle can provide more constraint on position estimates in the route direction. That is, their motion gradient is parallel to the route and, consequently, they constrain the vehicle location in the direction that most affects localization error. 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.

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 3 shows 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.

B. Region Selection

The question, now, is whether focusing on the regions containing more information actually leads to better localization. Without any knowledge as to which section of the image is more informative, an image descriptor would allocate equal weight to potentially informative regions (e.g., a building) as to regions with no information (e.g., the sky). Therefore, our hypothesis is that focusing feature extraction on the more informative image regions can lead to improved localization performance. We tested this hypothesis using the forward-looking camera from test route 2 (Figure 1d) and the localization algorithm described in Section IV to evaluate different sub-regions of the image. We sequentially removed the least informative regions of the image: the hood,
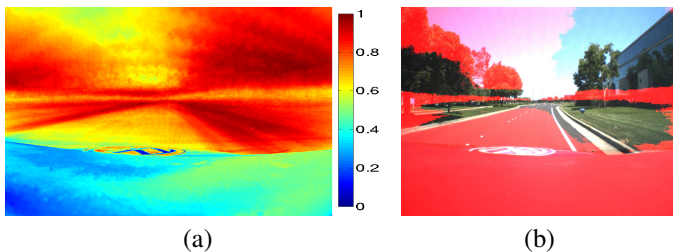

(a)   (b)

Fig. 3.   (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.
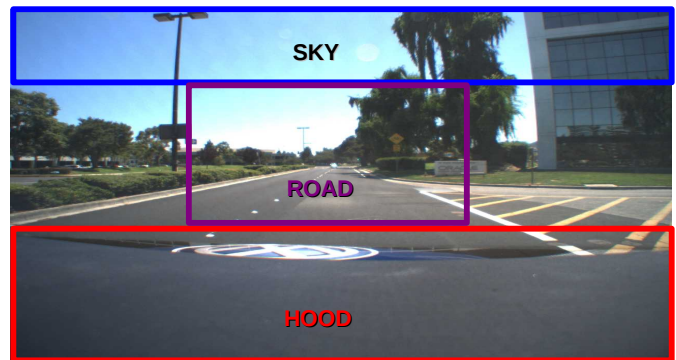

Fig. 4.   Region selection. Incrementally informative sub-regions: 1. Baseline (entire image); 2. Omit hood (red); 3. Omit hood/sky (blue); 4. Omit hood/sky/road (purple).

Fig. 5. Region selection using camera orientation = 45° and FOV = 90°.

	Error (m)			
Experiment Name	Avg	Max	Stddev	#Div
1. Baseline	3.58	20.43	4.04	0
2. Omit hood	1.39	12.65	2.11	0
3. Omit hood/sky	1.09	9.21	1.38	0
4. Omit hood/sky/road	0.75	6.45	0.83	0

	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)

the sky, and, finally, the road (Figure 4). The results, shown in Table II, show that localization accuracy increases when focusing on successively more informative sub-regions. We have also experimentally verified that the the uncertainty in the localization decreases. Selection of just the side-looking regions reduces the average error by 79%. Figure 2 shows graphically the effect of selecting informative sub-regions over an entire route.

To test the generality of these observations, we conducted a similar analysis using a different dataset that also used a different camera, optics, and route. In this case, we used images from test route 1, virtual cameras with 45° orientation from front-facing, and 90° FOV. Using manually selected regions (Figure 5) to discard the sky and road results in a reduction of localization error in every test case (Figure 6). Though the improvement in localization accuracy is less in this case, but the number of divergences is reduced to zero by region selection (Table III).

C. Camera Orientation

The previous section showed that informative regions of the image correlate well with localization accuracy. Now
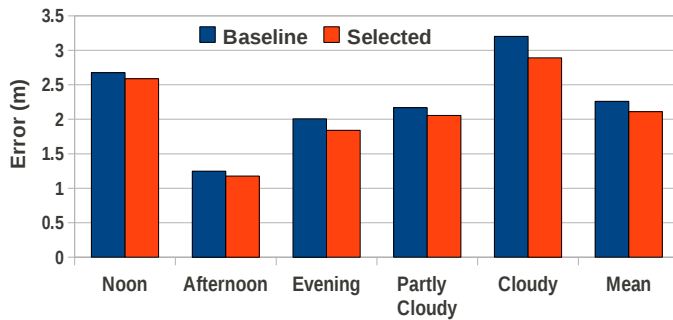
we return to the original question as to whether sideways-looking cameras improve localization accuracy, or, more precisely, which direction should a camera be pointed to achieve the best performance? To analyze this question, we used images extracted from the panoramic camera image, varying the virtual 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 7, 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).

D. Field of View

Field of view (FOV) is yet another parameter that can 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 IV, show that the best FOV choice depends somewhat on the camera orientation – a smaller FOV for the 45° orien-

Fig. 6. Result of using region selection shown in Figure 5 using the database from the "morning" trial.
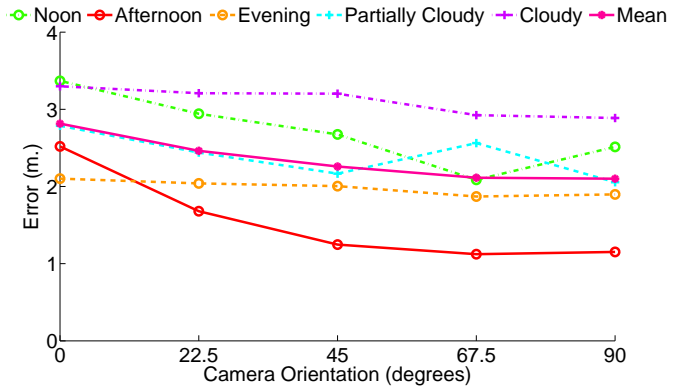

Fig. 7. The effect of camera orientation on localization accuracy.

TABLE IV

LOCALIZATION ERROR AND DIVERGENCES AS A FUNCTION OF CAMERA FOV FOR A GIVEN CAMERA ORIENTATION

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)

tation performed best, whereas the 75° FOV performed best on the 67.5° and 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 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.

E. 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 384x384 pixels down to 24x24 pixels. The results, shown in Figure 8 and Table V, show that reducing the image resolution does not have much of a detrimental effect on localization. Even down to a resolution of 24x24 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 [22]. Note that the good performance at low resolution may be tied to the image descriptor used for the localization.

F. 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.

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 9 and Table VI, 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 (35 m/s),
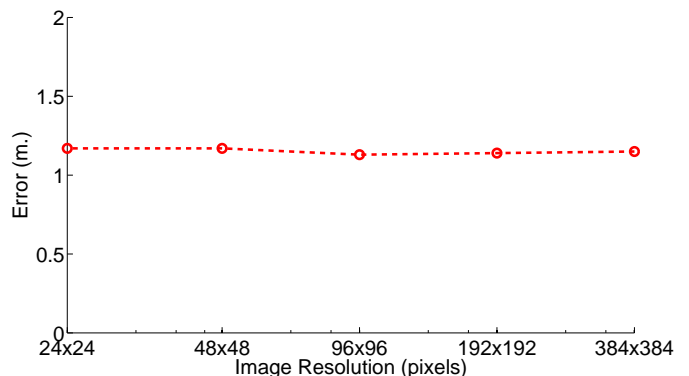
TABLE V

EFFECT OF IMAGE RESOLUTION ON LOCALIZATION ACCURACY.
CAMERA ORIENTATION = 67.5° AND FOV = 75°.

Image Resolution	Error (m)			
	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

Fig. 8.   Effect of image resolution

| Map Resolution | Error (m) | | | #Div |
	Avg	Max	Stddev	
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 VII

EFFECT OF ENVIRONMENT TYPE ON LOCALIZATION ACCURACY.

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

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.

VI. ENVIRONMENTAL FACTORS

The previous section focused on system 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 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 [6], [24].

A. 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.
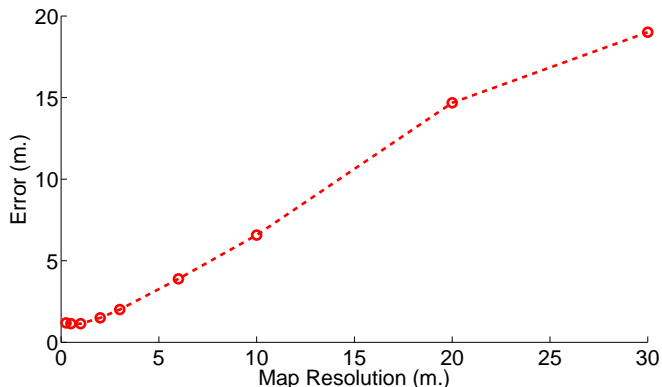
To analyze the effect of environment type on localization accuracy, we separated the data from test route 1 into three basic environmental types: rural/parklike, commercial/downtown, and residential/suburban. In fact, the test route was chosen to provide distinct, contiguous segments according to these categories. The results, shown in Table VII, 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 variability on the roadside due to changing parked cars in different trials. Another explanation may be that 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.

B. 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 7 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.

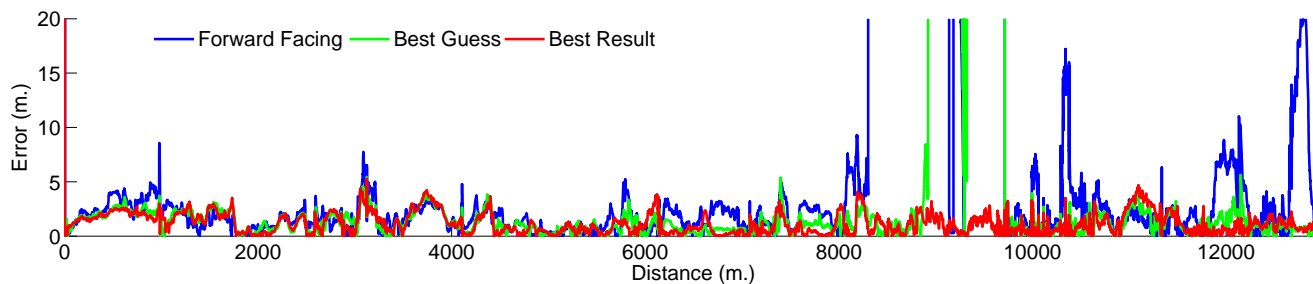Fig. 9.   Error as a function of map resolution.

Fig. 10. 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).

VII. Summary and Conclusions

In this paper, 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], [11], [12], [19]. 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 = $67.5°$, FOV = $75°$, image resolution = 192x192, map resolution = 1 m). The results (Figure 10) 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%. Finally, the raw data used in our analysis has been made publicly available for other researchers to use in benchmarking visual localization algorithms.

Acknowledgments

References

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

[2] M. Agrawal, K. Konolige, and M. R. Blas, "Censure: Center surround extremas for realtime feature detection and matching," in *ECCV*, vol. 5305, 2008, pp. 102–115.

[3] H. Andreasson, A. Treptow, and T. Duckett, "Localization for mobile robots using panoramic vision, local features and particle filter," in *ICRA*, 2005.

[4] H. Badino, A. Yamamoto, and T. Kanade, "Visual odometry by multi-frame feature integration," in *International Workshop on Computer Vision for Autonomous Driving*, 2013.

[5] H. Badino, D. F. Huber, and T. Kanade, "Visual topometric localization," in *Intelligent Vehicles Symposium*, 2011, pp. 794–799.

[6] ——, "Real-time topometric localization," in *ICRA*, 2012, pp. 1635–1642.

[7] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *Intl. Journal Robotics Research (IJRR)*, vol. 27, no. 6, 2008.

[8] M. Cummins and P. M. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *Intl. Journal Robotics Research (IJRR)*, vol. 30, no. 9, 2011.

[9] A. J. Davison, I. D. Reid, N. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, 2007.

[10] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Intl. Journal Robotics Research (IJRR)*, vol. 32, no. 11, 2013.

[11] 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.

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

[13] D. G. Lowe, S. Se, and J. J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *Intl. Journal Robotics Research (IJRR)*, vol. 21, no. 8, 2002.

[14] W. P. Maddern, M. Milford, and G. Wyeth, "CAT-SLAM: probabilistic localisation and mapping using a continuous appearance-based trajectory," *I. J. Robotic Res.*, vol. 31, no. 4, pp. 429–451, 2012.

[15] E. Menegatti, T. Maeda, and H. Ishiguro, "Image-based memory for robot navigation using properties of omnidirectional images," *Robotics and Autonomous Systems*, vol. 47, no. 4, 2004.

[16] M. Milford and G. Wyeth, "Mapping a suburb with a single camera using a biologically inspired SLAM system," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1038–1053, 2008.

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

[18] A. C. Murillo, G. Singh, J. Kosecká, and J. J. Guerrero, "Localization in urban environments using a panoramic gist descriptor," *IEEE Transactions on Robotics*, vol. 29, no. 1, pp. 146–160, 2013.

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

[20] D. Robertson and R. Cipolla, "An image-based system for urban navigation," in *IN BMVC*, 2004.

[21] G. Schindler, M. Brown, and R. Szeliski, "City-scale location recognition," in *CVPR*, 2007.

[22] 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.

[23] I. Ulrich and I. R. Nourbakhsh, "Appearance-based place recognition for topological localization," in *ICRA*, 2000.

[24] 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.