2015 IEEE International Conference on Robotics and Automation (ICRA)
Washington State Convention Center
Seattle, Washington, May 26-30, 2015

It's Not Easy Seeing Green: Lighting-Resistant Stereo Visual Teach & Repeat Using Color-Constant Images

Michael Paton, Kirk MacTavish, Chris J. Ostafew, and Timothy D. Barfoot

Abstract— **Stereo Visual Teach & Repeat (VT&R) is a system for long-range, autonomous route following in unstructured 3D environments. As this system relies on a passive sensor to localize, it is highly susceptible to changes in lighting conditions. Recent work in the optics community has provided a method to transform images collected from a three-channel passive sensor into color-constant images that are resistant to changes in outdoor lighting conditions. This paper presents a lighting-resistant VT&R system that uses experimentally trained color-constant images to autonomously navigate difficult outdoor terrain despite changes in lighting. We show through an extensive field trial that our algorithm is capable of autonomously following a 1km outdoor route spanning sandy/rocky terrain, grassland, and wooded areas. Using a single visual map created at midday, the route was autonomously repeated 26 times over a period of four days, from sunrise to sunset with an autonomy rate (by distance) of over 99.9%. These experiments show that a simple image transformation can extend the operation of VT&R from a few hours to multiple days.**

I. INTRODUCTION

Autonomous navigation in the absence of a Global Positioning System (GPS) over large distances and times is a crucial requirement for many potential robotic applications. To realize this goal, navigation algorithms need to rely exclusively on on-board sensors, scale to large environments, and cope with changes in the appearance of the world. Appearance-based localization and mapping techniques have allowed scalable autonomous navigation using only passive camera sensors, but have difficulty localizing as the appearance of the scene changes.

For example, Furgale and Barfoot developed Stereo Visual Teach & Repeat (VT&R) [6], an appearance-based navigation technique that allows a robot to autonomously repeat an arbitrarily long route using a stereo camera. This system exploits the fact that route following does not require a globally consistent state estimate; this task only requires a locally consistent estimate. The necessary map quality can be relaxed to being metrically consistent at the local level, and topologically consistent at the global level. The result is an algorithm that can navigate in large-scale environments.

To autonomously repeat a path, VT&R relies on extracting SURF [1] features from greyscale images, which are typically constructed using the green channel in an RGB camera. Extracted features are matched between consecutive images to obtain a motion estimate and matched between live and archived images to obtain pose estimates relative to the desired trajectory. This matching process quickly fails outdoors as the ambient lighting condition changes. As a result, autonomous visual (or appearance-based) navigation is severely limited outdoors. Although there have been efforts to counteract this issue [2], [12], it remains a problem for vision-in-the-loop navigation algorithms such as VT&R.

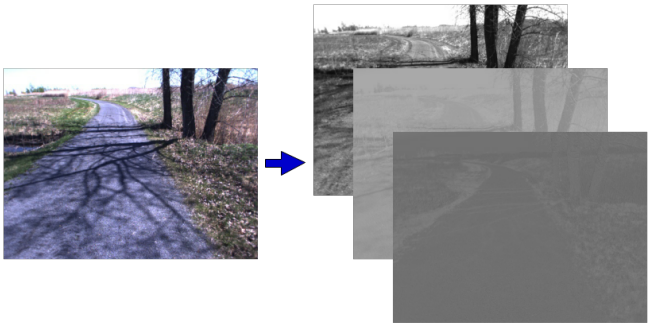


Fig. 1: This figure illustrates the transformation of an RGB image into a set of three greyscale images. The first is the image formed from the green channel, the second two are the experimentally found color-constant images (F_v, F_r) detailed in section IV. By making assumptions about the sensor and environment, a weighted log difference between channel responses can provide greyscale images that are invariant to changes in the ambient lighting conditions [14].

Inspired by recent developments in the research area of color constancy, this paper presents a stereo VT&R algorithm that is resistant to changes in lighting. Color constancy is the ability to perceive the color of objects as constant under varying illuminations. Recent publications have provided the means for computing color-constant greyscale images from a three-channel camera [5], [14], and have been shown to be successful in localization and mapping systems [3], [8]. We build on these works by using color-constant images, shown in Figure 1, in our stereo VT&R pipeline to achieve a vision-in-the-loop, autonomous route-following algorithm that is capable of handling significant lighting changes in a variety of outdoor environments.

Our algorithm's capability is demonstrated through an extensive field trial, where the robot repeats a 1km path over a period of four days from sunrise to sunset. This path spans harsh unstructured environments such as rocks and sand, windy grasslands, and strongly shadowed wooded areas. Throughout the trial we completed 26 repeats and maintained an autonomy rate of 99.9% by distance.

This paper is structured as follows. Section II describes recent related work. Section III presents the mathematical theory behind color-constant image transformations. Section IV overviews the training of our experimentally found color-constant images, as well as our lighting-resistant algorithm. Sections V and VI detail the field trial and its results, respectively. The experimentally found images are discussed in Section VII. The paper finishes with a conclusion.

978-1-4799-6923-4/15/$31.00 ©2015 IEEE

II. RELATED WORK

Lighting change in outdoor environments is a well-known problem for localization and mapping with passive sensors. The failure of appearance-based matching techniques in the face of lighting change restricts the usefulness of these algorithms. Recent research in the robotics community has attempted to overcome this issue.

An obvious way to deal with lighting is to switch to an active sensor. McManus and Barfoot [11] use a appearance-based LIDAR pipeline to perform lighting-invariant VT&R. Krúesi et al. [7] register dense 3D LIDAR scans with the Iterative Closest Point (ICP) algorithm to perform autonomous route following. Both algorithms are capable of autonomously repeating outdoor routes over a full 24 hour period. However, appearance-baed LIDAR is prone to motion distortion and dense registration suffers in open areas that lack geometric information.

Lighting invariance with a passive sensor is inherently more difficult. Churchill and Newman [2] build maps of parallel experiences to localize in environments that have predictable changes in appearance. Milford and Wyeth [12] align sequences of images to perform weak localization despite extreme changes in appearance, with the assumption that the robot's velocity is locally constant and the images are aligned [15]. Naseer et. al [13] compare sequences of images by building a directed acyclic graph of possible matches over the entire trajectory and a graph search to find the best match.

Appearance-based localization and mapping rely on consistent or predictable appearance. Color constancy is a feature of the human perceptual system that ensures that the perceived color of objects remains constant under varying lighting conditions. There is a wealth of optics research aimed at achieving color constancy with camera sensors. If assumptions are made about the sensor and the environment, a greyscale, lighting-invariant image can be obtained from a three-channel camera. Finlayson et al. [5] calculate a 2D colorspace that moves along a predictable direction as the lighting in the environment changes. By projecting the 2D colorspace onto a line that is orthogonal to this direction, a 1D colorspace that is invariant to lighting conditions is obtained. Ratnasignam et al. [14] extract a greyscale image by taking the weighted log difference of the three channels to cancel out the effects of illumination.

Recently, these optics theories have been tested by the computer vision and robotics community. Corke et al. [3] tested the image transformation described in [5] on a dataset of images captured under varying illumination conditions. They show an increase in precision/recall performance versus greyscale images when place recognition is performed on this dataset. Maddern et al. [9] localize monocular images against a prior map of colored 3D point clouds to obtain a 6-DoF pose estimate. By using color-constant images based on [14] during the day and greyscale images at night, they show successful localization over a 24-hour period. McManus et. al [10] run two separate localizers in parallel, one that uses greyscale images and one that uses color-constant images based on [14]. Localization with color-constant images only occurs when the greyscale localizer first fails. This work was shown to improve localization against a map that was collected in different lighting conditions, but was not employed within a vision-in-the-loop control system.

III. COLOR-CONSTANT IMAGES

A camera's response for a specific point, x, in the environment is described by the illuminant, sensor response, reflecting surface, and the geometry of the scene and camera. The light originates from an illuminant, is reflected by a surface towards the camera, and is focused onto an image sensor consisting of an array of filtered pixel sensors. This process results in the sensor response, R^x, describing the power of the light incident on the pixel sensor after being reflected and filtered. The illuminant is described by its intensity, I, and spectral power distribution, $E(\lambda,T)$, as a function of wavelength, λ, and temperature, T. At a specific point, x, the light is reflected according to the incident direction, $\underrightarrow{a}^x$, the surface normal, $\underrightarrow{n}^x$, and the surface reflectance, $S^x(\lambda)$. This light is filtered according to the sensor's channel, described by the spectral sensitivity, $F(\lambda)$. Integrating over the desired spectrum, ω, results in the image sensor response:

$$R^x = \underrightarrow{a} \cdot \underrightarrow{n}^x I \int_{\omega} S^x(\lambda) E(\lambda,T) F(\lambda) d\lambda. \quad (1)$$

Images that are resistant to the variation in the illumination of an outdoor scene can be calculated by making assumptions about the imaging sensor and environment [14]. These assumptions allow cancellation of the factors of (1) that are dependent on the scene's illumination: the spectral power distribution, $E(\lambda,T)$, and the intensity of the illuminant, I.

If the assumption is made that the spectral sensitivity function, $F(\lambda)$, is infinitely narrow at the sensor's peak waveform, λ_i, then (1) can be simplified to

$$R_i^x = \underrightarrow{a} \cdot \underrightarrow{n}^x I S^x(\lambda_i) E(\lambda_i,T). \quad (2)$$

Taking the logarithm of (2) will separate the surface reflectance, $S^x(\lambda_i)$, and geometric properties from the spectral power distribution of the light source, giving the following

$$\log(R_i^x) = \log(\underrightarrow{a} \cdot \underrightarrow{n}^x I) + \log(S^x(\lambda_i)) + \log(E(\lambda_i,T)). \quad (3)$$

Furthermore, if the assumption is made that the illuminant is a black-body radiator, then the spectral power distribution function, $E(\lambda,T)$, can be modeled using the Wein approximation [4], resulting in

$$E(\lambda_i,T) = 2hc^2 \lambda_i^{-5} e^{-\frac{hc}{k_B}/T\lambda_i}, \quad (4)$$

where T is the temperature of the black-body radiator, h is Plank's constant, k_B is the Boltzmann constant, and c is the speed of light. Substituting (4) back into (3) provides:

$$\log(R_i^x) = \log(\underrightarrow{a} \cdot \underrightarrow{n}^x I) + \log(S^x(\lambda_i) C_1 \lambda_i^{-5}) - \frac{C_2}{T\lambda_i}, \quad (5)$$

where $C_1 = 2hc^2$, and $C_2 = \frac{hc}{k_B}$. The result is a sensor response equation that separates the effect of illumination on the

scene from properties of the reflected surface material. A weighted linear combination of three channel responses can be constructed to effectively cancel out the first and third terms, providing an illumination-invariant sensor response that is affected only by the properties of the surface materials. This difference of log responses is provided on a per-pixel basis by the following equation [14]:

$$F = \log(R_2) - \alpha \log(R_1) + \beta \log(R_3), \tag{6}$$

where $\log(R_i)$ is (5) with peak waveform, λ_i, and weights α and β subject to the following constraints:

$$\frac{1}{\lambda_2} = \frac{\alpha}{\lambda_1} + \frac{\beta}{\lambda_3}, \quad \beta = (1-\alpha), \tag{7}$$

where $\lambda_1, \lambda_2, \lambda_3$ are the peak sensor responses ordered from lowest to highest wavelength. If these constraints are met, the weighted difference of the log responses will cancel out the effect of the spectral power distribution of the light source, $E(\lambda)$, and the illuminant intensity, $\underrightarrow{a}^x \cdot \underrightarrow{n}^x I$.

If an imaging sensor with at least four channels is provided, a second illumination-invariant image can be calculated. Combining the fourth channel with two of the others will provide the illumination-invariant feature F_2 [14]:

$$F_2 = \log(R_3) - \gamma \log(R_2) - \delta \log(R_4). \tag{8}$$

subject to the following constraints:

$$\frac{1}{\lambda_3} = \frac{\gamma}{\lambda_2} + \frac{\delta}{\lambda_4}, \quad \delta = (1-\gamma). \tag{9}$$

These equations provide the ability to calculate fast lighting-invariant greyscale images from cheap RGB sensors, providing a potential solution to the lighting problem associated with vision-based localization and mapping. This conversion is possible, however, due to two major assumptions made about the environment and the sensor (infinitely narrow sensor responses, black-body radiator illuminant). In the next section, we describe our experiments using this colorspace transformation and a stereo camera that significantly violates these assumptions, as well as our strategy to address the issue through training data.

IV. APPROACH

A. Static Experiments / Training the weights

All of our tests were conducted with a Point Grey Research Bumblebee XB3 stereo camera. This camera uses the Sony ICX445 CCD, whose photodetector sensor response is detailed in Figure 9. The response of each channel is spread out over a wide area with significant overlap at the peaks, violating the assumption of an infinitely narrow response.

Therefore, we decided to relax the second constraint introduced in (7). Doing so provided a free variable, α, with the weight, β, being calculated as follows:

$$\beta = \lambda_3 \left(\frac{1}{\lambda_2} - \frac{\alpha}{\lambda_1} \right), \tag{10}$$

which allowed us to experimentally tune the weights in (6) by analyzing the ability to match SURF features across significant lighting changes on collected training data. Relaxation of the second constraint in (7) no longer guarantees that the intensity of the illuminant with respect to the geometry of the scene is cancelled out, but allows us to find alternative color-constancy coefficients, (α, β), that perform well in practice.

By collecting timelapse stereo imagery of static outdoor scenes, and empirically tuning the color-constancy coefficients, we were able to improve SURF matches across drastic lighting changes for a given type of scene. We collected two separate datasets: a vegetation-heavy dataset, primarily residing around the green spectrum, and a rocks-and-sand dataset primarily residing around the red spectrum. Examples and results of these datasets are displayed in Figure 2.

Using these datasets, we noticed something interesting: the set of weights that performed well for the vegetation dataset performed poorly in the rocks/sand environment and vice versa. This led us to devise two complementary color-constant images based on the following equations:

$$\text{vegetation: } F_v = \log(R_2) - \alpha_v \log(R_1) + \beta_v \log(R_3), \tag{11}$$

$$\text{rocks/sand: } F_r = \log(R_2) - \alpha_r \log(R_1) + \beta_r \log(R_3), \tag{12}$$

where $\alpha_v = 0.29$, $\beta_v = 0.77$, $\alpha_r = -1.3$, and $\beta_r = 2.9$. Results from the static experiments are pictured in Figures 2(c) and 2(d). For each experiment, we analyzed the performance of the greyscale image, the fully-constrained, peak-invariant image, the F_v invariant image, and the F_r invariant image. The experiments show that SURF features are most resistant to changes in lighting using F_v in vegetation-heavy environments and F_r in rocks-and-sand. Motivated by the fact that our robots have no prior knowledge of their environment, we developed a lighting-resistant VT&R system that uses greyscale images, F_v images, and F_r images in parallel to achieve superior localization and Visual Odometry (VO).

B. Lighting-Resistant Stereo VT&R

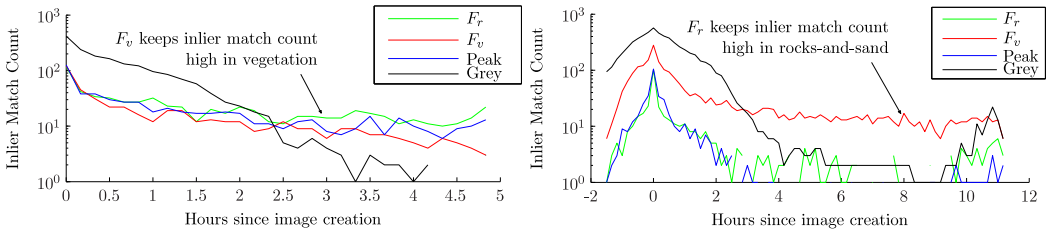
Stereo VT&R [6] allows a robot to autonomously follow a previously driven route using a stereo camera as the sole sensor. During the teaching phase, the robot is driven either manually or autonomously while a pose graph of visual features linked by relative transformations is built. The robot autonomously repeats the route by using VO and map localization to obtain pose estimates relative to the path. This information is fed to a path-tracking controller to keep the robot on the path.

1) Legacy System: The repeat-phase localization and VO pipeline for the legacy system [6] is shown and described in Figure 3. This pipeline describes how pose estimates relative to the map are extracted from the stereo image feed.

The map that is built during the teaching phase consists of keyframes of 3D SURF features, connected by relative transformations computed by the VO pipeline. To repeat the route, the map is searched until the live stereo feed localizes against the map. Upon localization, a current local submap consisting of the keyframe the robot is closest to, with a window of adjacent keyframes that are relaxed into a single coordinate frame, is loaded into memory.

At each iteration of a repeat, the robot performs VO and attempts to localize against the submap. The VO estimate is

(a) Vegetation Scene (b) Rocks-and-Sand Scene



(c) Vegetation Results (d) Rocks-and-Sand Results

Fig. 2: Static experiments. Overview with example photographs and SURF matching results for the experiments. Example images of each experiment are displayed in Figures (a) and (b). For each experiment, we empirically found a set of weights, (α, β), that generated a lighting-invariant image. This image provided superior results in terms of matching SURF features across lighting changes for the particular environment of the experiment. The weights found in the experiments are used to generate the lighting-resistant greyscale images, F_v and F_r. Results of the experiments are found in Figure (c) and (d) for the vegetation and rocks-and-sand experiments, respectively

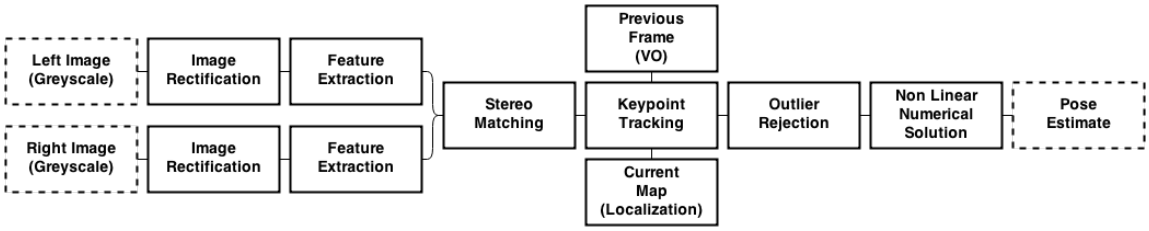


Fig. 3: Localization/VO pipeline for the legacy stereo VT&R system [6]. The input to the system is a left/right greyscale stereo image pair, the output is a pose estimate relative to a small subsection of the map. Incoming stereo images are first rectified. SURF visual features are then extracted from the images. Images are matched left-to-right to obtain depth for each found correspondence. These 3D keypoints are then matched to a small subsection of the map to obtain feature correspondences. Tracked keypoints are sent to an outlier rejection algorithm to filter out inliers and obtain an initial pose estimate. This pose estimate is used to prime a non linear numerical solver to obtain a refined pose estimate. VO estimates are obtained in the same fashion, with the exception that the live view is compared to the most recent view, instead of the map.

feed to the localizer as a prior, and is used as the final pose estimate if localization fails. Errors between the pose estimate and desired path are fed into a path-tracking controller to keep the robot in its tracks. Because this algorithm only relies on small relative submaps, it can scale to arbitrarily long maps, making it ideal for long-range outdoor autonomy. Due to the nature of appearance-based matching, however, VT&R's operational window is limited to a few hours after teaching in sunny conditions.

2) Lighting-Resistant System: Based on the results from the static experiments described in section IV-A, we have shown (empirically) that there exist two images that can be extracted from a three-channel camera that will increase the ability to match visual features across lighting changes in varying terrain. Despite the increase in robustness, it was shown by McManus et al. [10] that these images are noisier then their traditional greyscale counterpart. Greyscale images will provide a superior pose estimate as long as there are limited changes to lighting.

Using this intuition, and expanding slightly on the idea introduced by McManus et al. [10], our new algorithm combines the accuracy of greyscale images with the robustness of color-constant images to achieve superior localization. Our algorithm is identical to Furgale and Barfoot [6], with the

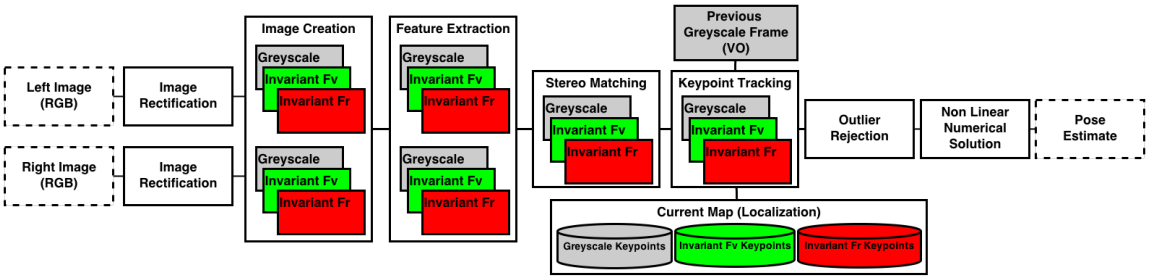


Fig. 4: Localization pipeline for the lighting-resistant stereo VT&R system. The input to the system is a left/right RGB stereo image pair, the output is a pose estimate relative to a small subsection of the map. Incoming stereo images are first rectified. The rectified RGB images are then converted to the three image sources: Greyscale, Invariant F_v, and Invariant F_r. SURF visual features are extracted from each image source independently. Visual features are matched left-to-right for each respective image source to obtain depth for each feature. These 3D keypoints are then matched to a small subsection of the map to obtain feature correspondences between the live view and the map. Tracked keypoints for each image source are placed together and sent to an outlier rejection algorithm to filter out inliers and obtain a single initial pose estimate. This pose estimate is used to prime a non linear numerical solver to obtain a refined pose estimate.

exception of the structure of the map and the localization pipeline. The localization pipeline is depicted and described in Figure 4. In the new pipeline, VO is performed using only the greyscale images, while localization is performed with all three. Matched features from all three image sources are fused to obtain a single pose estimate. Keyframes in the new map contain independent sets of SURF features from all three image sources, linked by relative transformations computed by VO.

In McManus et al. [10], matched features from one image source are chosen to obtain a pose estimate, favoring the greyscale matches. We found that by fusing matches from all image sources to obtain a pose estimate, we can achieve localization where matching from independent images would fail.

C. Localization Failure Conditions

In order for our system to repeat a route, it requires a pose estimate relative to the path. From this pose estimate, a path-tracking error can be fed into the path tracker, which aims to minimize this error. This pose estimate can be obtained from either localization to the current local sub-map, or a previous estimate propagated by VO. We consider a localization attempt successful when the system matches six or more visual features from the query image to the map after outlier detection. If matching to the map fails, then the pose estimate from the VO chain is used. Our system is typically capable of recovering after driving up to 20m using propagated VO (in the absence of map localizations). If the robot has driven over 20m without a successful localization, it is considered a localization failure, the robot stops, and the entire map is searched for a match to the current view.

V. VT&R Field Trial

A. Field Trial

We conducted an extensive field trial at the Canadian Space Agency (CSA)'s Mars Emulation Terrain (MET) at Montreal, Quebec. The MET is a 60m by 120m environment consisting of sand and rocks, emulating the surface of Mars.

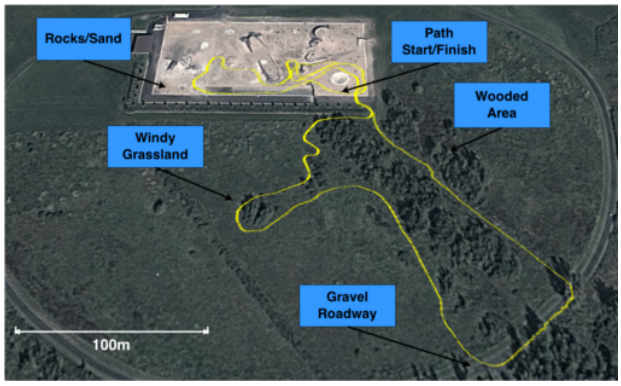


Fig. 5: Satellite imagery of the CSA Mars Emulation Terrain, with 1km teach route highlighted.

We chose the MET as an ideal testing environment for our algorithm due to the proximity of rock/sand and grass/forest regions, providing the possibility for a single route to contain both environments. To test our algorithm we taught a 1km route in sunny conditions and repeated the route 26 times over the course of four days, testing localization from sunrise to sunset.

The 1km route is displayed in Figure 5. It begins in the MET and travels approximately 300m through sand and rocks before entering into the adjacent field. The path then snakes through the field passing by grass, trees, and a gravel roadway. The path then passes through a wooded area, travelling alongside a stream. The path re-enters the MET, makes a short loop and finishes where the route begins. This route was taught at 10:50am on the first day of the trial during bright sunny conditions with strong shadows.

B. Hardware

The hardware configuration for the field trial is displayed in Figure 6. A Clearpath Robotics Grizzly RUV serves as our mobile robot platform. The Grizzly is equipped with a payload that includes a suite of interoceptive and exteroceptive sensors. For this field trial, the only sensor used for localization and mapping was the forward PGR Bumblebee

Fig. 6: Clearpath Grizzly RUV and its sensor configuration. The robot is equipped with a Point Grey Research Bumblebee XB3 camera, a GPS receiver, and a Hyundai generator. The robot contains a ROS enabled embedded computer that controls its motors and safety features. For this field trial, our algorithms ran on a Lenovo W540 laptop that interfaces with the on-board computer and the forward-facing Bumblebee XB3 camera.

XB3 stereo camera labeled in Figure 6. We collected GPS data during the route for the purpose of visualization only. All of our VT&R code ran on a Lenovo W540 laptop.

VI. RESULTS

The field trial proceeded by repeating the route approximately every hour from sunrise until sunset over two days, and one final time on the morning of the fourth day. This route was autonomously repeated 26 times, exposing the algorithm to significant differences in lighting conditions. Table I outlines our field test with autonomy rates. We analyzed the success of our system based on two metrics: the percentage of autonomous driving distance (vs. manual) for each repeat, and the distance the system needed to drive on VO without localization to the map. Through post-field analysis, we compare our results to the legacy VT&R localization system [6], as well as the state-of-the-art system that uses color-constant images [10].

A. Autonomy Rates/Manual Interventions

Our system is designed to autonomously repeat previously driven routes. In ideal situations, the robot will autonomously drive the entire distance. There are cases when manual intervention is required, however. If a localization failure described in section IV-C occurs, and re-localization has failed, then it is required to manually drive the robot a short distance to an area where it can relocalize. Also, if a collision with an obstacle is imminent, then a manual intervention to steer the robot to safety is required. During this field trial, only three of the repeats required manual interventions, totalling less than 0.1% of the 26km travelled.

TABLE I: Field trial schedule and autonomy rates.
*Note: *Autonomy Rates for this repeat are conservatively approximated due to a critical loss of data.*

Action	Day	Start	Summary	Autonomy
Teach	2014/05/12	1050	Sunny	—
Repeat 1	2014/05/12	1140	Sunny	100%
Repeat 2	2014/05/12	1253	Sunny	100%
Repeat 3	2014/05/12	1335	Sunny	100%
Repeat 4	2014/05/12	1400	Sunny	100%
Repeat 5	2014/05/12	1606	Sunny	100%
Repeat 6	2014/05/12	1727	Sunny	100%
Repeat 7	2014/05/12	1814	Sunny	99.2%
Repeat 8	2014/05/12	1900	Sunny	98.5%*
Repeat 9	2014/05/12	1929	Sunny	100%
Repeat 10	2014/05/12	2006	Sunset	100%
Repeat 11	2014/05/13	0620	Cloudy	100%
Repeat 12	2014/05/13	0705	Cloudy	100%
Repeat 13	2014/05/13	0800	Cloudy	100%
Repeat 14	2014/05/13	0900	Cloudy	100%
Repeat 15	2014/05/13	1000	Cloudy	100%
Repeat 16	2014/05/13	1100	Cloudy	100%
Repeat 17	2014/05/13	1200	Cloudy	100%
Repeat 18	2014/05/13	1300	Cloudy	100%
Repeat 19	2014/05/13	1400	Cloudy	100%
Repeat 20	2014/05/13	1510	Cloudy	100%
Repeat 21	2014/05/13	1600	Cloudy	100%
Repeat 22	2014/05/13	1700	Cloudy	100%
Repeat 23	2014/05/13	1800	Cloudy	100%
Repeat 24	2014/05/13	1900	Cloudy	100%
Repeat 25	2014/05/13	2000	Sunset	99.9%
Repeat 26	2014/05/13	2030	Dark	Failed
Repeat 27	2014/05/15	0850	Sunny	100%

The first day was sunny, allowing us to test the performance of our algorithm as the strong shadows changed. During this time, the robot autonomously repeated the route 10 times from 11:40 to 20:06. Of these repeats, eight were fully autonomous with only two requiring brief manual interventions. In both cases, the system was obtaining pose estimates from VO due to a lack of matches to the map. Intervention was required because the vehicle was slightly off its tracks and veering towards an obstacle. This is due to the inherent error drift in pose estimates from dead reckoning. Manual intervention consisted of slightly turning the robot back towards its tracks. In both cases, the robot had been driving on VO for less than the localization failure point, and continued to autonomously repeat after the intervention.

The second day was cloudy, allowing us to test our algorithm's performance as the intensity of the scene changed with no pronounced shadows. Fifteen repeats were completed from sunrise to sunset. Fourteen of these were 100% autonomous with the last requiring a small manual intervention to avoid a rock. It is worth noting that it was 30 minutes after sunset at the time of the intervention. The final attempt of the day was well past sunset and failed when the terrain was no longer illuminated by sunlight. This route was not included in the autonomy rate calculation as it was deemed too dark for any algorithm to work based on passive cameras.

The final day was sunny, allowing us to test the condition of sunrise with shadows. The final repeat occurred at 08:50 and was 100% autonomous.

B. Distance Driven on VO

The success of the system was primarily judged by analyzing how long the robot had to drive on dead reckoning without localizing to the map. Results from two repeats with respect to this criterion are presented in this paper. We chose these repeats for the following reasons: repeat 4 took place roughly three hours after map creation, this amount of time is typically when matching with greyscale images begins to fail. Repeat 7 contained a manual intervention, and occurred during the most difficult lighting condition to localize against: when shadows in the live view are oriented in the opposite direction of shadows in the map.

Figure 7 displays the Cumulative Distribution Function (CDF) of the distance the robot would have driven since a localization success for the entire traverse of both repeats using the following localization algorithms: Furgale [2010] [6] (grey), McManus [2014] [10] (red), and our algorithm (green). It reads as: "for Y% of the traverse, the robot drove less than X m on VO". Also displayed is the point when our system determines a localization failure (i.e., relied on pure VO for too long) and begins searching the map.

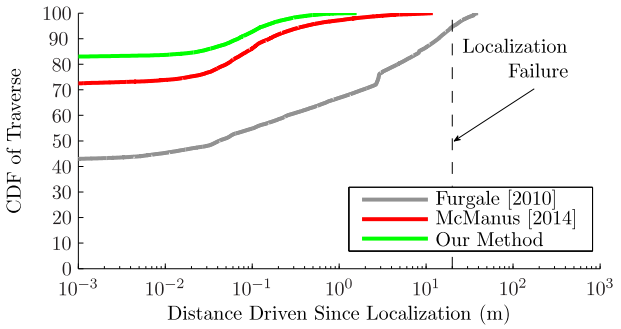
Localization based on greyscale image matching (Furgale [2010] [6]) would have resulted in driving on VO well beyond our system's localization failure point for both repeats. If color-constant images are used when greyscale fails (McManus [2014] [10]), the robot would have completed repeat 4, but would have had to drive over 39m on VO during repeat 7. Using our algorithm (green), distance driven without a localization never exceeded 2m during repeat 4 and 9m during repeat 7. We contribute the success of our system to two factors: the F_r image, and our method of fusing inlier matches from all three images to obtain a single pose estimate. The F_r image picks up extra matches in rocks and sand, and the fusing method provides localization success where matches from one image source would fail. An example of this is displayed in Figure 8.

VII. Discussion

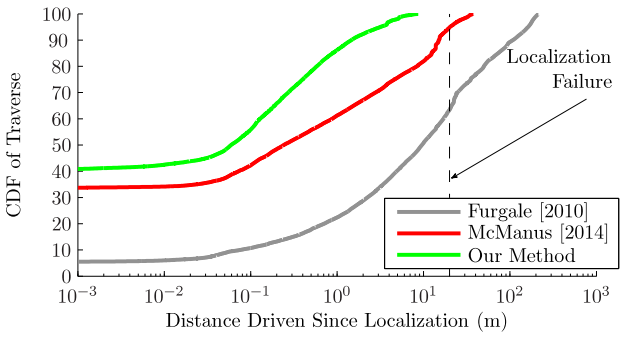
By breaking the constraints of (6) described in Section III, we have trained two color-constant images, F_v and F_r, that provide superior feature matching under large lighting changes. The weights used to generate these images are displayed in Table II.

The results of the experiments do not explain why breaking constraints would provide superior color-constant images, however. A closer look at the channel response of the RGB camera provides one potential explanation.

Figure 9 displays the sensor responses of a Sony ICX445 CCD, with the peak waveforms highlighted as $(\lambda_1, \lambda_2, \lambda_3)$. There is significant channel overlap at the peaks. To obtain color-constant images, the assumption that the camera's sensor responses are infinitely narrow centered at the peak wavelength must be made. The channel overlap at the peaks makes this assumption unrealistic for the given hardware and test environments. This may explain why an image formed from wavelength choices that are not centered at the peaks could provide superior results.



(a) Repeat 4 (3 hours after map creation))



(b) Repeat 7 (7.5 hours after map creation)

Fig. 7: Plot showing the CDF of the distance the vehicle had to travel on dead reckoning. Results are shown for the localization algorithms detailed in [6] and [10], as well as our method.

Fig. 8: Inlier feature tracks from the following image sources: greyscale (yellow), invariant F_v (green), and invariant F_r (red). Our system demands a minimum inlier match count of six features. Based on experience, any less results in a noisy pose estimate. *note:* additional results can be found in an attached video.

Table III lists weights for different wavelength choices that satisfy the constraints of Section III. Interestingly, if the choices for waveform peaks are made to reduce the overlap between channels, we arrive with weights that are very similar to the experimentally found weights of F_v. These waveform choices are displayed in Table III as F_v', and highlighted in Figure 9 as $(\lambda_1', \lambda_2', \lambda_3')$.

The F_r image outperformed both the peak image and F_v in the rocks-and-sand static experiment highlighted in section

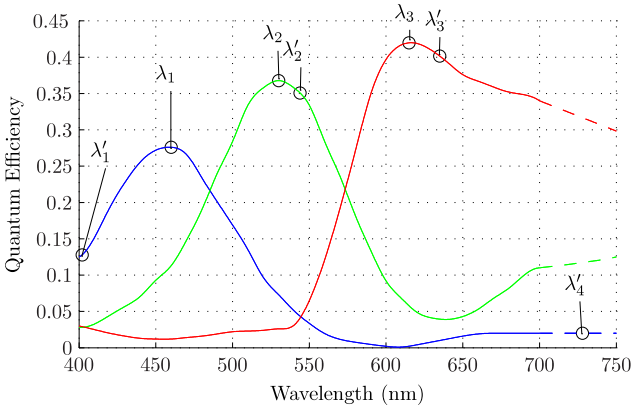


Fig. 9: Sensor response of the Sony ICX445 CCD, with wavelength choices highlighted (dashed lines are extrapolated). Peak wavelengths are denoted $(\lambda_1, \lambda_2, \lambda_3)$, Wavelengths approximating the weights used in our empirically tuned (F_v, F_r) are denoted $(\lambda'_1, \lambda'_2, \lambda'_3, \lambda'_4)$.

IV-A. An environment such as the rocks-and-sand static experiment or the surface of Mars will primarily reflect red light into an imaging sensor. This type of environment would effectively restrict the channel responses to the red/infrared wavelength range. Using this intuition and the sensor response, if the nominal wavelength choice for the blue channel is shifted to the red/infrared region of the sensor, we end up with weights that are very similar to F_r. The weights for F_r are very close to the weights displayed in Table III as F'_r. The waveform choices for F'_r are identical to F'_v, with the exception that the wavelength for the blue channel is 728nm. This waveform choice is shown as λ'_4 in Figure 9.

We conjecture that the wavelengths $(\lambda'_1, \lambda'_2, \lambda'_3, \lambda'_4)$ can be thought of as nominal choices for a 4-channel imaging sensor, where the upper portion of the blue channel response acts as a very weak fourth channel. These wavelength choices produce two color-constant images, (F'_v, F''_r), where F''_r is F'_r rescaled to conform to (8). We hypothesize that the experimentally found images (F_v, F_r) and their analogues (F'_v, F''_r) work well because the camera acts as a close approximation to a four-channel camera in the specific environments that we encountered. This would explain why F_v works well in typical green environments, and F_r works better in reddish rocks-and-sand environments.

TABLE II: Unconstrained ($\beta \neq 1-\alpha$) Experimental Weights

Image	λ'_1	λ'_2	λ'_3	λ'_4	α	β	γ	δ
F_v	460	530	615	—	.29	.77	—	—
F_r	460	530	615	—	-1.3	2.9	—	—

TABLE III: Constrained ($\beta = 1-\alpha$) Theoretical Weights

Image	λ'_1	λ'_2	λ'_3	λ'_4	α	β	γ	δ
Peak	460	530	615	—	.42	.58	—	—
F'_v	402	544	635	—	.29	.71	—	—
F'_r	728	544	635	—	-1.3	2.3	—	—
F''_r	—	544	635	728	—	—	0.45	.55

VIII. Conclusion / Future Work

Change in the lighting of a scene is a major problem for appearance-based localization algorithms that use passive sensors. We have presented an autonomous route-following algorithm based on color-constant image transformations that greatly increases the daily operational hours of the algorithm (i.e., from a few hours to the entire daylight envelope). We have experimentally validated this claim through an extensive field trial, where a robot repeated over 26 $\times$ 1km of unstructured terrain with an autonomy rate of 99.9% despite significant changes in lighting (sunrise to sunset).

We are currently investigating the utility of this new color-constant pipeline in the Canadian winter, where low sun elevation and snow provide extreme conditions for visual navigation.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) through the NSERC Canadian Field Robotics Network (NCFRN).

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359, 2008.

[2] W.S. Churchill and P. Newman. Experience-based navigation for long-term localisation. *Int. Journal of Robotics Research*, 2013.

[3] P. Corke, R. Paul, W. Churchill, and P. Newman. Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. In *Intelligent Robots and Systems (IROS)*, Nov. 2013.

[4] G.D. Finlayson and S.D. Hordley. Color constancy at a pixel. *J. Opt. Soc. Am. A*, 18(2):253–264, Feb 2001.

[5] G.D. Finlayson, S.D. Hordley, L. Cheng, and M.S. Drew. On the removal of shadows from images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(1):59–68, Jan 2006.

[6] P. Furgale and T.D. Barfoot. Visual teach and repeat for long-range rover autonomy. *Journal of Field Robotics*, 27(5):534–560, 2010.

[7] P. Krüsi, B. Bücheler, F. Pomerleau, U. Schwesinger, R. Siegwart, and P. Furgale. Lighting-invariant adaptive route following using icp. *Journal of Field Robotics*, 2014.

[8] W. Maddern, A. Stewart, C. McManus, B. Upcroft, W. Churchill, and P. Newman. Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles. In *Proc. of the Visual Place Recognition in Changing Environments Workshop,Robotics and Automation (ICRA)*, May 2014.

[9] W. Maddern, A. Stewart, and P. Newman. LAPS-II: 6-DoF day and night visual localisation with prior 3D structure for autonomous road vehicles. In *IEEE Intelligent Vehicles Symposium (IV)*, June 2014.

[10] C. McManus, W. Churchill, W. Maddern, A. Stewart, and Paul Newman. Shady dealings: Robust, long- term visual localisation using illumination invariance. In *In Robotics and Automation (ICRA)*, 2014.

[11] C. McManus, P.T. Furgale, B.E. Stenning, and T.D. Barfoot. Visual teach and repeat using appearance-based lidar. In *Robotics and Automation (ICRA)*, 2012.

[12] M.J. Milford and G.F. Wyeth. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *Robotics and Automation (ICRA), 2012 IEEE Int. Conf. on*, 2012.

[13] T. Naseer, L. Spinello, W. Burgard, and C. Stachniss. Robust visual robot localization across seasons using network flows. In *AAAI*, 2014.

[14] S. Ratnasingam and S. Collins. Study of the photodetector characteristics of a camera for color constancy in natural scenes. *J. Opt. Soc. Am. A*, 27(2):286–294, Feb 2010.

[15] N. Sünderhauf, P. Neubert, and P. Protzel. Are we there yet? challenging SeqSLAM on a 3000 km journey across all four seasons. In *Proc. of the Workshop on Long-Term Autonomy, Robotics and Automation (ICRA)*, 2013.