Taming the North: Multi-Camera Parallel Tracking and Mapping in Snow-Laden Environments

Arun Das, Devinder Kumar, Abdelhamid El Bably, and Steven L. Waslander.

Abstract - Robot deployment in open snow-covered environments poses challenges to existing vision-based localization and mapping methods. Limited field of view and over-exposure in regions where snow is present leads to difficulty identifying and tracking features in the environment. The wide variation in scene depth and relative visual saliency of points on the horizon results in clustered features with poor depth estimates, as well as the failure of typical keyframe selection metrics to produce reliable bundle adjustment results. In this work, we propose the use of and two extensions to Multi-Camera Parallel Tracking and Mapping (MCPTAM) to improve localization performance in snow-laden environments. First, we define a snow segmentation method and snow-specific image filtering to enhance detectability of local features on the snow surface. Then, we define a feature entropy reduction metric for keyframe selection that leads to reduced map sizes while maintaining localization accuracy. Both refinements are demonstrated on a snow-laden outdoor dataset collected with a wide field-of-view, three camera cluster on a ground rover platform.

1 Introduction

A wide range of challenging and remote tasks have been proposed as possible field robotics applications, from wilderness search and rescue, to pipeline and infrastructure inspection, to environmental monitoring. Particularly in Northern climates, these activities require autonomous navigation in snow-laden environments, which present distinct perception challenges for autonomous vehicles. The possibility of tree cover precludes reliance on GPS alone for positioning, and both obstacle detection and accuracy requirements further drive the need for alternate localization methods.

Both visual and laser based simultaneous localization and mapping methods can provide such improved localization. Although laser scanners are not significantly affected by snow, their relatively large costs can be prohibitive for many applications. In this work, we consider the problem of deploying a feature based visual SLAM system known as Multi-Camera Parallel Tracking and Mapping (MCPTAM) in a snowy, outdoor environment. MCPTAM employs an arbitrary cluster of cameras with wide field of view, with or without overlap, to track point features in the envi-

University of Waterloo. Waterloo, Ontario, Canada
e-mail: {arun.das, devinder.kumar, ahelbably, stevenw }@uwaterloo.ca

ronment, and has been demonstrated to provide accuracy better than 1% of distance traveled in both indoor and outdoor environments [17, 8, 7].

The primary challenge with outdoor and snowy environments is that large areas of the image are relatively feature poor due to limited geometric structure, overcast skies and large regions of uniform snow cover. Without employing expensive high dynamic range cameras, this leads to difficulties tracking features near the robot and clusters the points used for map generation along the horizon. The result is poor translational tracking and a susceptibility to map optimization failures if features are incorrectly corresponded.

To address these limitations, we introduce two extensions to our previous work. First we investigate changes to MCPTAM's front-end, by pre-processing the camera frames to extract more robust features. We use as motivation some of the works of [9, 19] which use region based contrast equalization and horizon detection [20] to fulfill this goal. Second, we propose core changes to MCPTAM's backend which allow for more informed keyframe selection based on the expected entropy reduction of uncertainty in the map points. These modifications directly impact the quality of the localization solution by creating a more robust set of features to track and optimize against for mapping.

2 Related Works

To date, there have been comparatively few instances of autonomous robotic deployments in snowy conditions. The CoolRobot is a mobile sensor station deployed both in Greenland and on the Antarctic plateau, and relies on solar power and GPS waypoint navigation to move through primarily flat terrain [13]. Similarly, both the Nomad [1] and MARVIN [6] rely on GPS guided navigation with a laser scanner and vision for local collision avoidance in polar environments. The SnoMote platform seeks to augment GPS with visual localization and terrain drivability estimation for detailed ice sheet mapping [19].

Closely related to visual navigation in snow-covered terrain is use of computer vision for planetary exploration. The visual localization challenges are similar in both environments, with limited local features, large variations in scene depth, and unreliable features in the sky portion of images. For example, stereo localization has been used on lengthy datasets collected in Devon Island, Canada [5], where repetitive ground terrain and a lack of rotation invariant features led the authors to note the concentration of features on the horizon. Similarly, stereo and/or laser scan data was employed in a large range of planetary analog terrains for localization and drivability analysis [18]. In both cases, the image quality both near the robot and at a distance was not often an issue for feature extraction.

The MCPTAM method builds on the foundation of Parallel Tracking and Mapping [10], which splits the localization and mapping problem into separate pose tracking and keyframe based feature mapping processes. This divide prevents pose estimation from being delayed by the batch optimization required as a part of the mapping bundle adjustment. Features are tracked between images and localization

is performed relative to the known map, while map updates are performed when new keyframes are selected to be inserted into the global map.

Many visual mapping techniques use keyframes in order to reduce the computational burden of the mapping process. Existing approaches generally insert keyframes based on point triangulation baseline [10], or other heuristics such as the co-visibility of features [16], or the overlap in the number of tracked points [12]. These heuristics attempt to insert keyframes in order to maintain the map integrity, yet do not directly attempt to minimize the uncertainty in the map. The work most related to ours generates image features off-line, creates a buffer of the image frames, and selects keyframes based on saliency in order to reduce content redundancy [3]. In contrast, our approach is a real-time, online system, and attempts to reduce feature uncertainty while the camera is in motion.

In addition to keyframe selection, the identification of strong and stable visual features is both important and challenging in snowy environments. The Snomote [19] integrates a pre-processing technique of contrast limited adaptive histogram equalization (CLAHE) to enhance the contrast of the captured images. A slope finding method is applied to mask out the mountain peaks or other structures from the background and SIFT features are detected mainly from the foreground.

Applying feature detection methods to the entire image is problematic, however, as environments with trees and foliage result in self similar image features which are difficult to match. Instead, horizon detection can be used to apply specific feature detection criteria in the snow-laden region of the image. Existing methods (e.g. [4]) do not explicitly consider the snow-laden case, with the exception of the SnoMote [20], which uses a weighted sum of weak and strong visual cues to identify fairly precise horizon lines. The method is overly computationally expensive for our application, and so we present a simplified method based on the Hough transform in this work.

3 Multiple Camera Parallel Tracking and Mapping

MCPTAM is a real-time, feature-based, visual slam algorithm which extends Klein and Murray's Parallel Tracking and Mapping (PTAM) [10] in five ways. First it allows multiple, non-overlapping field-of-view (FOV), heterogeneous cameras in any fixed configuration to be successfully combined. MCPTAM's novel initialization mechanism allows for scale to be recovered, even with non-overlapping cameras. Second it extends the PTAM's pinhole camera model to work with fish-eye and omnidirectional lenses through the use of the Taylor camera model [15]. The ultra-wide FOV coupled with the multi-camera cluster prevents feature starvation due to occlusions and textureless frames in any single camera. Third, PTAM's backend has been replaced with the g2o optimizer allowing for faster and more flexible optimization structures [11]. Finally, MCPTAM introduces both an improved update process based on box-plus manifolds and a novel feature parameterization using spherical co-ordinates anchored in a base-frame [17].

A brief overview of the MCPTAM formulation proceeds as follows. Denote a point in the global frame, $p \in \mathbb{R}^3$ as $p = [p_x\ p_y\ p_z]^T$ where p_x , p_y , p_z represent the x, y, and z components of the point, respectively. Let the map, P, be a set of points,

defined as $P = \{p_1, p_2, \ldots, p_n\}$. Denote the re-projection function as $\Pi : \mathbb{R}^3 \mapsto \mathbb{R}^2$, which maps a point in the global 3D frame to a pixel location on the image plane.

In the standard pinhole camera model, light rays are represented as lines which converge at the center of projection and intersect with the image plane. In order to accommodate the large radial distortion caused by fisheye lenses, the Taylor model uses a spherical mapping where the elevation and azimuth angles to a 3D point, $s = [\theta, \phi]^T$, are modeled as half lines which pass through the sphere's center, which are then mapped to the image plane through a polynomial mapping function.

In order to track the camera cluster pose, $\omega^c \in \mathbb{SE}(3)$, the map points, P, are reprojected into the image frames of the cameras. Given a set of feature correspondences, the camera cluster pose parameters are found through a weighted nonlinear least squares optimization which seeks to determine the pose parameters such that the re-projection error between corresponding points is minimized. By re-observing features, the point locations in the map can be refined using additional measurements, and new map points can be inserted into the map. To perform these tasks, MCPTAM uses *keyframes*, which are a snapshot of the images and point measurements taken from a point along the camera cluster's trajectory. Since MCPTAM performs tracking using multiple cameras, it extends the idea of key-frames to *multi-keyframes*, which are simply a collection of the key-frames from the individual cameras at a particular instant in time.

We shall define a multi-keyframe, M, as collection of keyframes, $M = \{K_1, \ldots, K_m\}$, corresponding to the m individual cameras which are part of the multi-camera cluster. Each multi-keyframe is associated with its pose in $\mathbb{SE}(3)$. In order to insert a new multi-keyframe into the map, the point measurements from each observing keyframe are collected, and the parameters of the point locations, as well as the keyframe poses are optimized using a bundle adjustment procedure.

Entropy Computation for a Gaussian PDF: The Shannon entropy is a measure of the unpredictability or uncertainty of information content. Suppose $X = \{x_1, x_2, \ldots, x_n\}$ is a discrete random variable. The Shannon entropy for X, $H(X)$ is given as $H(X) = -\sum_{x_i \in X} P(x_i) \log P(x_i)$, where $P(x_i)$ denotes the probability of event x_i occurring. The Shannon entropy provides a scalar value that quantifies the average variance of the discrete random variable X. The base of the logarithm denotes the units of the entropy. In the case where the base of the logarithm is 2, the units are referred to as *bits*, and when performed using the natural logarithm, the units are referred to as *nats*. It is also possible to compute the Shannon entropy for a continuous random variable. In the case where the continuous random variable is modeled as a Gaussian distribution, the entropy can be computed as

$$h_e(Y) = \frac{1}{2} \ln((2\pi e)^n |\Sigma|), \tag{1}$$

where Σ is the covariance matrix of the multivariate Gaussian distribution, $|\cdot|$ denotes the determinant operator, and $h_e(Y)$ is used to denote that the logarithm was taken with base e. Note that unlike the entropy for discrete random variables, it is possible for the entropy of continuous random variables to be less than zero.

4 Proposed Approach

Our approach involves both pre-processing of images to improve feature tracking despite the limitations of images acquired in snow-covered environments, and improvements to the keyframe selection process that help maintain map quality throughout the test datasets.

4.1 Pre-Processing Pipeline

The pre-processing pipeline that is used to enhance the captured image for detecting good features for localization of our mobile robot consists of snow segmentation, histogram equalization, and feature selection phases.

Snow Segmentation: We first apply a Canny edge detector [2] to remove the undesired information from the image while still retaining the structural information. This is applied prior to a Hough Line transform, which is used to detect the line that segments out the snow from the rest of the regions in image.

Consider a line represented in the polar form $\rho = x\cos\theta + y\sin\theta$ where ρ is the radial distance from the origin and θ is the angle formed by this radial line and the horizontal axis measured in the counter-clockwise direction. The Hough Line transform uses a 2D accumulator array to detect the existence of lines in the edge based image from the Canny edge detector using a voting based method to output ρ and θ. Each element, (ρ, θ), in the output represents a line. For our task , we select the element with the highest value as the horizon, which indicates the straight line that is the most strongly represented in the input image. It is important to note that for our concerned task, we only detect horizontal lines in the image.

Histogram Equalization: Before feeding the input image to MCPTAM we use histogram equalization to enhance the global contrast of the image. Since snow laden environments lead to low contrast images, enhancing the contrast can significantly improve the detection of stable features. The global histogram equalization (GHE) transform, $T(r)$, can be represented as

$$T(r) = (L-1)\sum_{j=0}^{L-1} p_r(r_j), \tag{2}$$

where L represents the number of gray level intensities present in the image, j is the intensity level varying from 0 to $L-1$, and $p_r(r_j)$ is the probability distribution function (pdf) of intensity level j.

The pdf is defined by:

$$p_r(r_j) = \frac{N_j}{N_t}, \tag{3}$$

where N_j is the number of pixels with intensity level j and N_t is the total number of pixels present in the image. We also implemented contrast limited adaptive histogram equalization (CLAHE) [21] for comparison. Instead of accounting for global illumination changes and coming up with single histogram, CLAHE computes several histograms each belonging to a different part of the image and uses

this information for changing the *local* contrast of the image. CLAHE also contains a contrast limiting function that limits the amplification of noise.

Feature Selection: We take this enhanced image obtained after histogram equalization and input it into MCPTAM system where we detect coarse, mid level and fine FAST features in the images for each camera. FAST features are used because of their computational efficiency and ability to detect stable corner features [14]. Using the (ρ, θ) obtained from the Hough Line transform, we select fine features from the segmented snow region below the horizon, and coarse features from the rest of the image. The large structural features in the snow laden environments are generally trees or far away buildings, and generating fine features from these image regions are not helpful as the features generated are not sufficiently distinguishable to produce correct correspondences. The nearby features in snow on the ground can be better localized, however, and therefore become very important to the mapping process. Hence we detect and track fine features in snow and coarse features from far away structures for localization and mapping.

4.2 Entropy Based Keyframe Selection

The quality of the map point parameter estimation is heavily dependent on the triangulation baseline between the measurement viewpoints. Many visual SLAM techniques use heuristics based on the point triangulation baseline to perform keyframe insertion, however no existing approaches attempt to perform keyframe selection through direct minimization of the point estimate covariance.

We propose a covariance update on the point with the assumption that the keyframe candidate's location is known and fixed. Although the keyframe's pose parameters are in fact updated through bundle adjustment once inserted into the map, the fixed keyframe parameter assumption allows for rapid evaluation of the point covariance update, and is reasonable so long as the tracker pose estimate is sufficiently accurate.

In order to determine when a multi-keyframe should be inserted into the map, we inspect the uncertainty of the current camera cluster provided by the tracking process. The covariance of the tracking pose parameters is given by $\Sigma^c = (G^T W G)^{-1}$, where $G = \frac{\partial \Pi}{\partial \omega^c}$ is the Jacobian of the map re-projection error with respect to the cluster state, and W is the matrix of weights associated with the measurements. To assess the current tracking performance, we extract the x, y, and z diagonal components of covariance matrix Σ^c, denoted as σ_x, σ_y, σ_z, respectively. The rotational covariances are ignored at this stage, as generally the rotations of the camera cluster can be tracked accurately using points of varying depth, whereas accurate positional tracking requires relatively close points in order to resolve the scale of the motion. Finally, a multi-keyframe is added when any element of the positional entropy is above a user defined threshold, ε, or

$$\max(h_e(\sigma_x), h_e(\sigma_y), h_e(\sigma_z)) > \varepsilon, \tag{4}$$

where $h_e(\cdot)$ is computed using Equation (1). When a multi-keyframe addition is triggered, the next step is to determine which multi-keyframe should be added. For

this, multi-keyframe candidates are maintained in a buffer and scored based on the expected reduction in point depth entropy if added to the map through a bundle adjustment process.

As the tracking thread operates, each successfully tracked frame, along with its corresponding set of point feature measurements and global pose estimate, are added as multi-keyframe candidates in a buffer. Suppose the tracking thread is currently operating at time t, and the last multi-keyframe insertion occurred at time k. Denote the set of multi-keyframe candidates which are buffered between times t and k as

$$\Phi = \{M_t, M_{t-1}, M_{t-2} \ldots, M_{t-k}\}. \tag{5}$$

Since each of the multi-keyframe candidates are saved from the tracking thread, an estimate of the global pose of each candidate is available from the tracking solution. Therefore, it is possible to determine the subset of map points observed in the individual keyframes within each multi-keyframe candidate. Denote the set of map points from P, visible in $K_l \in M_i$, as $\tilde{P}_{K_{il}} \subset P$.

Since each map point position is estimated through a standard bundle adjustment approach, the map point parameters are modeled as a Gaussian distribution with an associated mean and covariance. We denote the estimate for point p_j as $\hat{p}_j$, and the associated covariance matrix $\Sigma_j \in \mathbb{R}^{3\times 3}$.

Suppose point $p_j \in \tilde{P}_{K_{il}}$ is observed in keyframe $K_l \in M_i$. Our method seeks to determine the updated covariance of point p_j, if triangulated using an additional measurement from keyframe K_l. This is accomplished using a covariance update step as per the Extended Kalman Filter.

Denote the Jacobian of the re-projection function with respect to the point parameters, p, evaluated at point $\hat{p}_j$, as

$$J_j = \frac{\partial \Pi}{\partial p}|_{\hat{p}_j}. \tag{6}$$

The Jacobian, J_j, describes how perturbations in the point parameters for $\hat{p}_j$ map to perturbations in the image re-projections. Using the Jacobian, J_j, and the prior point covariance Σ_j, the predicted point covariance is given as

$$\bar{\Sigma}_j = (I - \Sigma_j J_j^T (J_j \Sigma_j J_j^T + R)^{-1} J_j)\Sigma_j. \tag{7}$$

The predicted covariance $\bar{\Sigma}_j$ provides an estimate of the covariance for point p_j, if the observing keyframe was inserted into the bundle adjustment process. Note that Equation (7) can be evaluated rapidly for each point, as the computational bottleneck is the inversion of a 3 by 3 matrix.

Although comparison of the predicted covariance to the prior covariance provides information on reduction of point parameter uncertainty for one point, the covariance representation does not allow for a convenient way to asses the uncertainty reduction across all of the points observed in the multi-keyframe. To that end, we propose evaluation of the uncertainty reduction using the point *entropy*.

Denote the entropy corresponding to the point's prior and predicted covariance as $h_e(\hat{p}_j)$ and $\bar{h}_e(\hat{p}_j)$, respectively. The reduction in entropy for point p_j is given as $\Lambda(p_j) = h_e(\hat{p}_j) - \bar{h}_e(\hat{p}_j)$. Using the expected entropy reduction for a single point, the expected entropy reduction for all of the points observed in multi-keyframe M_i is given as $\Psi(M_i) = \sum_{K_l \in M_i} \sum_{p \in \tilde{P}_{K_{il}}} \Lambda(p)$. Finally, when a multi-keyframe needs to be inserted into the map, all of the multi-keyframes within the buffer, Φ, are evaluated for total point entropy reduction. The multi-keyframe selected for insertion, M_i^*, is the one from the buffer which maximizes the point entropy reduction:

$$M_i^* = \underset{M_i \in \Phi}{\operatorname{argmax}} \, \Psi(M_i). \tag{8}$$

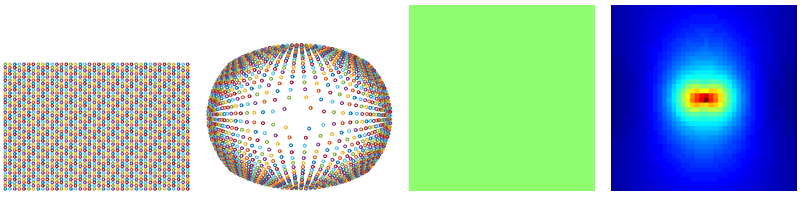
Once the optimal keyframe from the buffer is selected, it is inserted into the map through bundle adjustment, and the multi-keyframe buffer, Φ, is cleared.

Although it is possible perform keyframe selection using heuristics which rely on the geometric relationships between point observation baselines, such approaches do not account for possible degradation of point re-projection sensitivity that is also dependent on the camera model. For example, an image taken from a wide field of view fisheye lens camera will generally have significant distortion and spatial compression near the image edges. To illustrate this point, consider a uniform, 2D, planar grid of points, positioned at unit depth from a camera. Figures 1(a) and 1(b) show the projection of the grid onto the image plane using the pinhole and Taylor models, respectively. The pinhole projection preserves the uniform spatial distribution of the 3D grid on the image plane, while the Taylor model spatially compresses the points near the boundaries of the image plane. Such compression suggests that with a large FOV lens described using the Taylor camera model, the point projections which fall near the boundaries of the image are less sensitive to perturbations of the 3D point location. This insight is illustrated in Figures 1(c) and 1(d), which show the norm of the projection Jacobian with respect to perturbations in the x direction of the 3D point grid. It is evident that the pinhole camera model maintains uniform sensitivity to point perturbations across the image plane, while the Taylor camera model has reduced sensitivity as the points are projected farther from the image center.

Our proposed keyframe selection method is able to account for the properties of the lens model being used, as the point projection Jacobian, given by Equation (6), is dependent on the underlying camera model. For example, using the Taylor model, Equation (6) can be expanded as

$$\frac{\partial \Pi}{\partial p} = \frac{\partial \Pi}{\partial s} \frac{\partial s}{\partial r} \frac{\partial r}{\partial p} \tag{9}$$

where $r \in \mathbb{R}^3$ is the position of point p with respect to the observing frame, $\frac{\partial \Pi}{\partial s}$ relates the image re-projection to the point's projection on the unit sphere, $\frac{\partial s}{\partial r}$ relates the perturbations of a point projection on the unit sphere to perturbations of the point position in the observing keyframe, and $\frac{\partial r}{\partial p}$ relates the changes of the point in the observing keyframe to changes of the point parameters.



(a) Pinhole (b) Taylor (c) Jacobian (pinhole) (d) Jacobian (Taylor)

Fig. 1 Comparison of image re-projection sensitivity between pinhole and Taylor camera models. (a) and (b) illustrate the projection of 3D points onto the image plane, using the pinhole and Taylor camera models, respectively. The image compression around the edges results in reduced sensitivity of image projection Jacobian in the outer edge areas, as seen in (d), where as the pinhole camera model displays uniform strength in the image re-projection Jacobian, as seen in (c).

5 Experimental Results

To verify our proposed methods, experiments were conducted using field data collected in a snow laden environment. A Clearpath Robotics Husky platform was equipped with three Ximia xiQ cameras, arranged in a rigid cluster, with one camera looking forwards, and the others facing off to the left and right sides of the vehicle. The cameras were fitted with wide angle lenses, with approximately 160 degrees field of view. Images were captured at 30 frames per second, at a resolution of 900x600 pixels. The vehicle traveled at a constant velocity of 0.5 m/s for over 120 m, and traversed a snow and ice covered path, as well as a snowy field area.

5.1 Image Pre-processing

We compare GHE and CLAHE in terms of the features that result after pre-processing. The FAST features detected on snow in the enhanced images are shown in Figure 2. It is evident that the largest number of features detected in snow were found with GHE. To quantitatively compare the two histogram equalization techniques, we calculated the number of features detected below the horizon. The total number of features obtained for a video sequence of 1497 frames from our dataset were 407,665 for GHE, 83,650 for CLAHE and 4,919 without any histogram equalization, demonstrating the advantage of GHE in terms of FAST feature detection in snow.

Fig. 2 Comparison of FAST features detection on (a) a normal image, (b) image ehanced by global histogram equalization, (c) image ehanced by CLAHE.

For segmenting the snow from the rest of the image, representative results are shown in Figure 3, which includes the output of our snow segmentation algorithm (the red line) for the single frontal view (camera 1) and features detected on snow in Canny edge images for the three camera cluster. Our approach produces a rough segmentation of each image in 0.015s, on average, over the entire dataset, which has an image resolution of 900x600. To compare our approach with the the state of the art result [20], we decrease the resolution of our captured dataset to 640x480. A naive implementation of our approach took on average 0.0098 seconds per frame, whereas the method proposed in [20] requires 0.0296 seconds per frame.

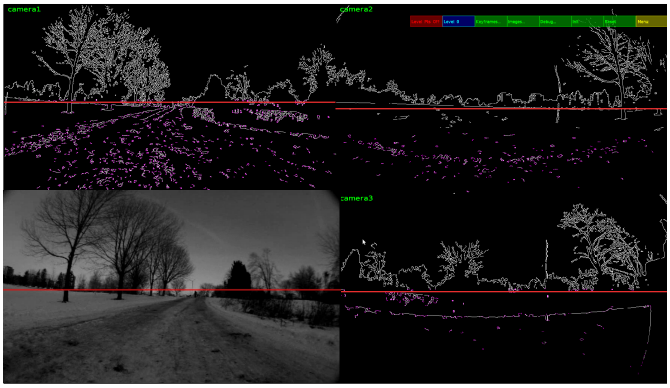
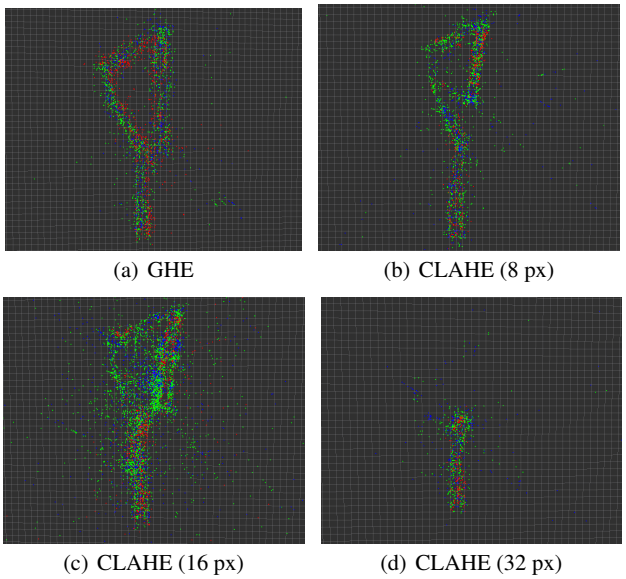


Fig. 3 The result of snow segmentation from camera 1 (lower left) and FAST features detected on snow in Canny edge images for the three cameras.

5.2 MCPTAM using Histogram Equalization

We next compare MCPTAM mapping performance with different equalization methods. As evident in Figure 4, GHE provides the most consistent feature map, compared to the CLAHE methods. As the patch size for the CLAHE methods increase, the resulting map exhibits signs of scale drift, as well as poor feature matches. It is also worth noting that GHE results in the recovery of a greater number of fine features, compared to the adaptive method. This is likely because GHE maintains more consistent illumination between the inserted keyframes, resulting in better feature matches over local methods.

Table 1 presents a summary of the results. It is evident that GHE resulted in a feature map with the fewest number of inserted multi-keyframes and the fewest number of points. This suggests the robot was able to travel longer distances on average before inserting a multi-keyframe into the map and localize more accurately with the features that were included, which is further verified by the reported maximum tracking entropy over the trail. The GHE method resulted in the lowest tracking entropy (as calculated by Equation (4)), suggesting the generated map provided stable points to track against throughout the test run.



(a) GHE (b) CLAHE (8 px)

(c) CLAHE (16 px) (d) CLAHE (32 px)

Fig. 4 Comparison of feature maps with different histogram equalization techniques. Red points denote fine features, while blue and green points denote coarse features. (a) shows the resulting map when the images are processed using GHE. Figures (b) to (d) present maps generated using CLAHE with different patch sizes. Note that large patch sizes cause instability in the feature tracking due to mismatched points.

Table 1 Summary of Results for Histogram Equalization Experiments

	GHE	**CLAHE (8)**	**CLAHE (16)**	**CLAHE (32)**
Max. Tracking Entropy (nats)	-2.4851	-2.2738	-1.5941	-1.6170
No. Map Points	2777	2960	6115	-
No. MKFs	168	175	240	-

5.3 Multi-keyframe Selection

Although previous authors have successfully used keyframe insertion methods related to feature overlap and the number of features tracked, such approaches were completely unsuccessful for our application due to intermittent feature tracking experienced in snowy environments. Instead, we compare our entropy based (EB) approach to a movement threshold on the vehicle, where a multi-keyframe is inserted once the camera cluster moves a user defined threshold distance from the previously inserted multi-keyframe. Only a threshold on the position is used; the rotation need not be considered due to the nearly 360 degree view of the multi-camera cluster, which tends to maintain consistent orientation based on stable, persistent horizon features.

Figure 6 presents a comparison of the multi-keyframe selection methods tested. The EB approach provides consistent mapping results, while the 2m threshold ap-

proach fails midway through the path. This is likely because the non-entropy based approaches do not consider any improvements in the map points, and merely assume that the multi-keyframe insertion will improve the map and provide stable points to track against. Our approach, on the other hand, actively seeks to insert multi-keyframes such that the map integrity is maintained, providing the camera cluster with stable and well estimated point features for localization. Although the map generated by the 1m threshold (1mt) policy (Figure 6(a)) is qualitatively similar to the one generated by the EB approach, the 1mt map contains approximately 42% more points compared to our proposed method, as summarized in Table 2. From Table 2, it is also clear that the EB method results in the lowest tracking entropy, along with the fewest inserted multi-keyframes. This is because our approach only adds new multi-keyframes when required by the tracker, and seeks to improve the points which exist in the map. As a result, fewer multi-keyframes are added, and fewer points are required to maintain suitable tracking integrity.

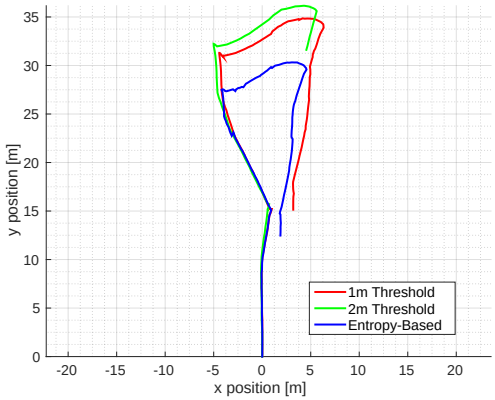
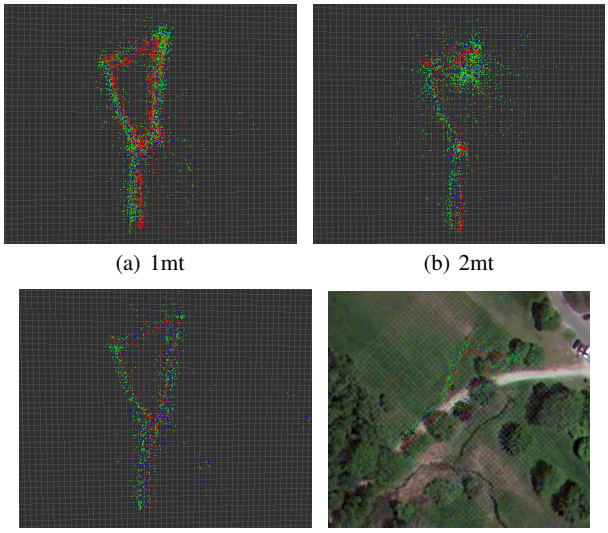


Fig. 5 Comparison of the recovered vehicle motion using different multi-keyframe selection methods. Note that the EB approach demonstrates the lowest scale drift in the trajectory.

Figure 5 presents the recovered vehicle trajectories. As seen in Figure 6(d), the vehicle traverses along a path area, then moves onto a field, and finally joins up with the path again. All of the evaluated methods result in similar trajectories over the path area, but exhibit differences once the vehicle moves onto the field. We see that the EB multi-keyframe selection approach results in the smallest scale drift while traversing the field, as demonstrated by the path closely rejoining itself. Conversely, the 1mt and 2mt approaches both exhibit a larger scale drift in the trajectory solution, since the static threshold policies do not account for map integrity when inserting multi-keyframes.

6 Conclusion

In this work, two extensions to the MCPTAM visual localization method are shown to significantly improve the performance of the system in snow laden environments.



(a) 1mt (b) 2mt

(c) EB (d) Map Overlay

Fig. 6 Comparison of multi-keyframe selection methods. Figures (a) and (b) show the resulting map using a 1m and 2m movement threshold, respectively. Figures (c) and (d) present the generated map using our proposed entropy based keyframe selection method.

Table 2 Summary of Results for multi-keyframe selection experiments

	EB-MKF	**1m Threshold**	**2m Threshold**
Max. Tracking Entropy (nats)	-2.76771	-2.0314	-2.4113
No. Map Points	2316	4001	2897
No. MKFs	150	175	162

We demonstrate that a pre-processing pipeline that uses GHE to improve FAST feature detection in snow, as well as horizon detection and a tailored feature selection process, results in improved feature tracking. We also show that point entropy reduction can be used as a keyframe selection metric, which leads to fewer keyframes and reduced map drift when compared to existing methods. In the future, we intend to expand the set of environments employed for testing, incorporate ground truth measurement of vehicle motion, and investigate the persistence and accurate localization of features in the map.

References

1. Apostolopoulos, D.S., Wagner, M.D., Shamah, B.N., Pedersen, L., Shillcutt, K., Whittaker, W.L.: Technology and field demonstration of robotic search for antarctic meteorites. The International Journal of Robotics Research **19**(11), 1015–1032 (2000)

2. Canny, J.: A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on **8**(6), 679–698 (1986)
3. Dong, Z., Zhang, G., Jia, J., Bao, H.: Keyframe-based real-time camera tracking. In: Computer Vision, 2009 IEEE 12th International Conference on, pp. 1538–1545. IEEE (2009)
4. Ettinger, S.M., Nechyba, M.C., Ifju, P.G., Waszak, M.: Towards flight autonomy: Vision-based horizon detection for micro air vehicles. In: Florida Conference on Recent Advances in Robotics, vol. 2002 (2002)
5. Furgale, P., Barfoot, T.D.: Visual teach and repeat for long-range rover autonomy. Journal of Field Robotics **27**, 534560 (2010)
6. Gifford, C.M., Akers, E.L., Stansbury, R.S., Agah, A.: Mobile robots for polar remote sensing. In: The Path to Autonomous Robots, pp. 1–22. Springer (2009)
7. Harmat, A., Sharf, I., Trentini, M.: Parallel tracking and mapping with multiple cameras on an unmanned aerial vehicle. In: Proceedings of the International Conference on Intelligent Robotics and Applications, vol. 1, pp. 421–432. Montreal, QC (2012)
8. Harmat, A., Trentini, M., Sharf, I.: Multi-camera tracking and mapping for unmanned aerial vehicles in unstructured environments. Journal of Intelligent & Robotic Systems pp. 1–27 (2014)
9. Kim, A., Eustice, R.M.: Real-time visual slam for autonomous underwater hull inspection using visual saliency. IEEE Transactions on Robotics **29**(3), 719–733 (2013)
10. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), pp. 225–234 (2007)
11. Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: g2o: A general framework for graph optimization. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (2011)
12. Leutenegger, S., Furgale, P.T., Rabaud, V., Chli, M., Konolige, K., Siegwart, R.: Keyframe-based visual-inertial slam using nonlinear optimization. In: Robotics: Science and Systems (2013)
13. Ray, L.E., Lever, J.H., Streeter, A.D., Price, A.D.: Design and power management of a solar-powered cool robot for polar instrument networks. Journal of Field Robotics **24**(7), 581–599 (2007). DOI 10.1002/rob.20163. URL http://dx.doi.org/10.1002/rob.20163
14. Rosten, E., Drummond, T.: Fusing points and lines for high performance tracking. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, vol. 2, pp. 1508–1515. IEEE (2005)
15. Scaramuzza, D., Martinelli, A., Siegwart, R.: A flexible technique for accurate omnidirectional camera calibration and structure from motion. In: IEEE International Conference on Computer Vision Systems (ICVS), pp. 45–45. IEEE (2006)
16. Stalbaum, J., Song, J.B.: Keyframe and inlier selection for visual slam. In: Ubiquitous Robots and Ambient Intelligence (URAI), 2013 10th International Conference on, pp. 391–396 (2013)
17. Tribou Michael J. and. Harmat, A., Wang, D., Sharf, I., Waslander, S.L.: Multi-camera parallel tracking and mapping with non-overlapping fields of view. International Journal of Robotics Research: to appear pp. 1–43 (2014)
18. Wettergreen, D., Wagner, M.: Developing a framework for reliable autonomous surface mobility. In: International Symposium on Artificial Intelligence, Robotics, and Automation in Space (iSAIRAS) (2012)
19. Williams, S., Howard, A.M.: Developing monocular visual pose estimation for arctic environments. Journal of Field Robotics **27**(2), 145–157 (2010)
20. Williams, S., Howard, A.M.: Horizon line estimation in glacial environments using multiple visual cues. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 5887–5892. IEEE (2011)
21. Zuiderveld, K.: Contrast limited adaptive histogram equalization. In: P.S. Heckbert (ed.) Graphics Gems IV, pp. 474–485. Academic Press Professional, San Diego, CA (1994)