Stereo Algorithm Selection using Region Based Similarity Measure for Reconstruction of Sea Ice Images

Anonymous WACV submission

Paper ID ****

Abstract

It is known that disparity maps are generated by matching similar features in stereo image pairs. The quality of the stereo analysis depends on the quality of these features. Because ice images have very low texture, it is challenging to find and match such features. In this work, we hypothesize that a group of pixels in an image originates from different but unknown distributions. Gaussian Mixture Model on different correspondence metrics is used to identify and classify the aforementioned distributions. Multiple algorithms are tested to determine which produces better analysis for these groups. We analyze for the most effective stereo measures however, none of those individually are capable of performing the stereo analysis of ice images accurately. For each stereo pair, four state-of-the-art stereo generating algorithms from our study are finally applied. The winning algorithm is recorded for each group based on the least RMS error. The results show that our selective stereo matching is better than just using one algorithm for the entire stereo pair.

1. Introduction

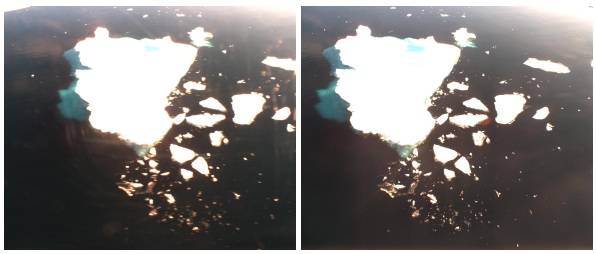
Sea ice keeps the polar regions cool and helps moderate global climate. It has a bright surface because eighty percent of the sunlight that strikes it is reflected back into space. The melting of sea ice uncovers darker land or ocean beneath, which then absorbs more sunlight [1], causing more heat to be absorbed leading to global warming [2, 3]. Melting sea ice also exposes certain regions of earth to big waves that can cause erosion. In the past decades [4], the sea ice cover has melted considerably and that continues to pose a problem.

The primary objective of this work is to perform stereo analysis on sea ice and get an estimate of the ice roughness and thickness in some regions. The Polar Sea Ice Topography Reconstruction System (PSITRES) [5] stereo camera system had been developed for this research. Eventually the goal is to be able to automate the process of using the image data from certain regions of sea ice cruises and perform the analysis over time to quantify the changes. The fully developed stereo systems could be deployed on sea ice cruise ships and could automatically acquire data over regular periods of time to perform the task. The advantage would be to have a system that can objectively perform the analysis for large regions of sea ice whereas manual measurement is cumbersome.

In this work, we use the previously collected data to perform stereo analysis of the sea ice in order to get an estimate of the thickness in the northern polar regions. There are some large stereo analysis data sets, such as the full-size 2005 Middlebury stereo pairs [6] with average of 1.4 megapixels resolution having a disparity range of 200 pixels, but the recent Disney/ETH datasets [7] are as large as 19 megapixels with disparity ranges up to 1000 pixels. While some methods try to perform stereo analysis in short time using approximations [8], some others perform the same more accurately but having longer run times [9].

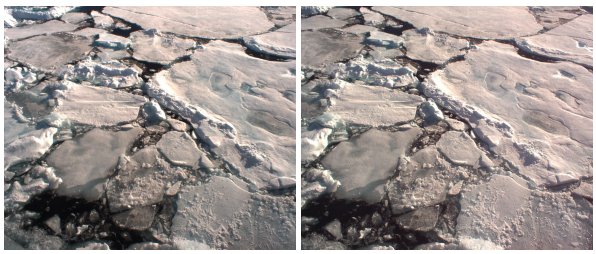
Historically, various algorithms have been proposed for reconstructing 3D scenes from stereo image pairs. These algorithms rely on the accurate stereo correspondences for disparity estimation. The disparity estimation becomes difficult in presence of occlusions, object boundaries, and fine structures, which can appear blurred in one or both stereo images. Stereo analysis for local algorithms is performed by selecting the disparity with the lowest matching cost, that is, winner takes all [10]. In a stereo analysis pipeline [11], the first step is to calibrate the stereo camera pair, which involves recording a series of checkerboard patterns with varying poses at a specific distance from the stereo rig. The image pairs, typically between ten and twenty pairs, are then processed via various software packages to derive the stereo camera calibration parameters.

Figure 1 shows a pair of stereo images from the ice



(a) (b)

Figure 1: Image pair with less features and high exposure (a) Left Image from Oden Arctic Technology Research Cruise (OATRC) (b) Right Image from OATRC



(a) (b)

Figure 2: Image pair with more features due to fragmented ice (a) Left Image from OATRC (b) Right Image from OATRC

database of the Fram Straight and the Greenland Sea in 2013 aboard the Oden as part of the Oden Arctic Technology Research Cruise (OATRC). The first stereo pair illustrates the fact that it is challenging to extract feature points from the data due to low texture and high reflection of light. The second pair of images shown in Figure 2 is preferred over the first for extracting rectification parameters due to the fact that it offers more corners which is evident from the ice fragments. This image pair also comes from the same expedition, but taken on a different day.

The next step in the stereo pipeline is that of rectification of the image pairs, using the parameters captured above. Rectification is the process of transforming or warping each image (of the pair) such that search for feature correspondences is simplified. Once rectified, the search is reduced from the two-dimensional search problem to only one dimension, improving the efficiency of the search.

The process of rectification can be performed in two ways. One can use the stereo camera calibration parameters that are obtained during calibration of the stereo rig, or by using the image pairs, choosing matching feature points and calculating the transformation needed to warp the images accordingly. The calibrated method failed for our images, because the calibration parameters never worked on ice images.

Feature extraction in images where regions have extremely low texture becomes challenging, which is exacerbated by the unavailability of ground truth data. Manual annotation of the ice regions is a challenging, error prone and time consuming work. It is also impractical for large-scale and real-time uses. Hence we use a semi-guided method to extract feature points which is described later. The matching of two regions or patches for similarity has been explored in several works [12, 13, 14, 15, 16]. Deep learning methods have been proposed that learn the matching likelihood of image patches with better generalization properties [17].

As mentioned earlier, image pairs similar to Figure 2 are preferred for deriving the rectification model in this work. This is primarily due to the availability of discernible corners for the algorithms to discover feature points. Hence, SIFT (Scale Invariant Feature Transform) [18] features have been used with an implementation by VLFeat.org to detect and match features on the ice images and then verified for accuracy. In the absence of ground truth each image pair needs to be carefully examined to verify the rectification process, which requires that the same features in the rectified images will have the same Y-coordinate.

Disparity, which is the shift of a feature point on the right image from its corresponding match on the left image, is inversely proportional to the depth of the feature in the scene. Before rectification, the image pair is converted to grey-scale for ease of computation, because only feature points matter, not their color information. This is due to the fact that the ice images are primarily white to various shades of grey and rarely have any color variation. The surrounding water, if present in the image, appears very dark as in Figure 1.

Various algorithms perform stereo analysis defined by different requirements. Sometimes accuracy is compromised to achieve fast computation and vice-versa. For certain applications like autonomous driving, both speed of computation and accuracy are crucial [19]. End-to-end deep neural networks can achieve great accuracy in performing specialized work, but may not be desirable in general cases [20, 21, 22, 23]. More methods need to be explored for accurate stereo analysis of ice images. In this work, multiple methods have been tested on the ice data after preprocessing the image pairs. Our method has the following steps:

- STEP 1: Find and verify point correspondences using SIFT between stereo image pairs, generating sample of ground truth for each stereo pair
- STEP 2: Create varying size kernels around these correspondences
- STEP 3: Evaluate the underlying distribution that these kernels may have been generated from, using different metrics and Gaussian Mixture Models
- STEP 4: Classify regions of the image pair into groups using the resulting distributions
- STEP 5: Empirically determine which method works best for each group, and using the one that performed the best for that group of pixels

Below we discuss few methods that are closely related to this work, however we note that none of these alone can produce accurate disparity map for ice images.

2. Related Work

Hirschmuller [24] generates disparity map by considering pixel-wise matching of Mutual Information and approximating a global, 2D smoothness constraint by combining many 1D constraints. Matching cost is calculated hierarchically by using Mutual Information. The metric is insensitive to recording and illumination changes. It is defined from the entropy H of two images (that is, their information content), as well as their joint entropy. Cost aggregation is the result of approximation of a global energy function by path-wise optimizations from all directions through the image. Disparity computation is done by winner takes all (WTA) and supported by disparity refinements like consistency checking and sub-pixel interpolation.

Sinha et al. [25] try to estimate the final disparity by first selecting multiple (top-m) candidates per pixel, instead of just selecting the surface with the minimum aggregated cost. At each pixel, they compute the median of all multi-valued disparity hypotheses within a T x T window centered on that pixel and select this as the final disparity. To select the multiple candidates, they first compute the minimum cost c at a pixel, and then select the top-m planes that have costs less than or equal to lambda times c. For a pixel with fewer than m candidates, they replicated its winning candidate to ensure that it has m hypotheses. They have used value of m as 2 and lambda as 1.25. Tau has been set to 5 for resolutions of 3 megapixel or less, 7 otherwise.

Kim et al. [26] first perform reliable stereo analysis specifically around object boundaries instead of interior regions, by operating on individual light rays instead of image patches. More homogeneous interior regions are then processed in a fine-to-coarse procedure rather than the standard coarse-to-fine approaches. Global optimization is not performed at any point in their method in any form. That allows the algorithm to retain precise object contours while still ensuring smooth reconstructions in less detailed areas. The image space disparity is defined for a pair of images captured at adjacent positions or, which is the displacement between the two adjacent horizontal lines in an Epipolar Plane Image. We do not use this approach because in this work we only focus on RGB images of relatively lower resolution rather than the very high resolution images with 3D light fields.

Yoon et al. [7] in their work assumes that the disparity varies smoothly in each region of the varying window size, according to the classic Gestalt grouping theory. Further, they assume that neighboring disparities, if corresponding to the same object in a scene, are similar and that two neighboring pixels with similar disparities support each other. They compute the support-weights of the pixels in a given support window using color similarity and geometric proximity, in accordance with Gestalt principles. As a result, their method does not depend on the initial disparity estimation because the adaptive support-weight computation is entirely based on the contextual information within a given support window. The matching cost between pixels is measured by aggregating raw matching costs with the support weights in both reference and target support windows. The computed dissimilarity can be erroneous when considering only the reference support window, in case the target support window has pixels from different depths. To minimize the effect of such pixels, they compute the dissimilarity between pixels by combining the support-weights in both support windows. The combined support-weights favor the points likely to have similar disparities with the centered pixels in both images. After the dissimilarity computation, the disparity of each pixel is simply selected by Winner-Takes-All method without any global reasoning. However, these methods require precise color segmentation that is very difficult to achieve for the scope of this work.

Batsos et al. [27] in their work generate a matching volume leveraging data with ground truth and conventional wisdom to perform stereo analysis. They proceed by coalescing diverse evidence from a bidirectional matching process via random forest classifiers. Then they show that the resulting matching volume estimation method achieves similar accuracy to purely data-driven alternatives on benchmarks and that it generalizes to unseen data much better. We do not have ground truth values for each pixel of

the stereo image pairs and hence unable to test this method.

It is worth mentioning that Li et al. [9] use segment-based method that can tackle large texture-less regions without losing performance. They present a multiple proposal scheme: Point-wise competition picking the best absolute disparities which effectively suppresses outliers. Pairwise collaboration that casts a vote on relative disparities that can distinguish the artificial boundaries. The proposals are used as priors within a unified Markov Random Fields model. Using left and right rectified image pair, they calculate the initial disparity using off-the shelf method. Then they obtain a set of 3D surface fragments from the disparity map and 2D segments, which is derived from the M-segmentation of the reference left image. The 3D surface fragments generate absolute disparities with risks and relative proposals. Proposals with pixels falling outside of the surface fragments are rejected. The output disparity map is the one that fits the global model best using a variant of the widely used semi-global matching.

Lastly, the following method has been used by Cech et al. [19] to visit only a small fraction of the disparity space to find a semi-dense disparity map by growing a small set of correspondence seeds. The method can recover from wrong initial assignments, affecting only the computation time and not the quality of the final stereo analysis. The disparity components are kept growing regardless of their overlap in disparity step, temporarily foregoing the uniqueness constraint in order to use random seeds. The growth process is stopped at loci where the final optimal result cannot be improved, ensuring visiting a small fraction of the disparity space. Then the global optimality task is solved by a robust matching algorithm that selects among the competing components in disparity space. The matching algorithm solves a graph-theoretic problem of maximum strict sub-kernel which has a strict stability quality. Instead of stopping the growth whenever the uniqueness condition is violated, it is stopped when the image similarities of the overlapping disparity components differ by more than a set threshold. This strategy primarily helps in developing the stereo analysis that is capable of accurate matching of repetitive or uniform patterns like the ice images, as demonstrated in the results.

Most of these works have either used a single method to perform stereo analysis or have limiting assumptions and have achieved good results in a specific task. Some have generalized the domain and unable to produce accurate map for ice images. Instead of using hand-crafted rules for the matching process, many authors have used classifiers [28, 29, 30, 31, 32, 33, 34]. In this work, lack of ground truth and texture on ice create challenges that have not been addressed before. Instead of using one method for the entire stereo image, multiple methods have been tested and used for grouped regions where each method performs better than the others. This has resulted in performing stereo analysis that is more accurate than that resulting from using one method alone on the entire stereo image pair.

3. Understanding and Pre-processing Data Sets

3.1. The Ice Database

The ice database is composed of data collected from three separate research expeditions in ice covered waters [5]. These expeditions were completed aboard three separate research vessels in different parts of the Arctic and at different times of year. The PSITRES camera system was first deployed in 2012 aboard the RV Polarstern for 80 days over a large region of the central Arctic, as well as the Berents, Kara, and Laptev seas. In total PSITRES has spent 118 days at sea, collecting 8,048,715 images, which resulted in 4.5 TB of high quality data. It has been re-calibrated on board for every expedition to maintain high quality of data collected. Each image from the ice database is 2448 x 2048 pixels, acquired by a pair of 5 MP FLIR Point Grey camera housed in Dotworkz Fire Enclosures mounted on the flying deck of the ice going vessels. The cameras use 8 mm focal length Computar lenses with a 1.75 meter baseline and have a viewing area of 1000 square meter. The cameras are triggered with a hardware switch at approximately 2 FPS. Ethernet cable is used to transfer data between the cameras and the ship's on-board computer.

3.2. Data Pre-processing

As is well known in computer vision, rectification can be done either via calibrated or uncalibrated route. Because the calibrated route failed on ice images, the uncalibrated route was used. In the absence of recorded ground truth, attempt was made to hand select matching points on the image pairs but proved to be not of good enough quality for the transformations that were calculated based on those. Hand clicking less than thousand points resulted in severely distorted transformations which could not be used for rectification and disparity computations.

Widely available feature detection techniques like Harris Corner and SURF features failed to detect matches between left and right stereo pairs on ice data due to lack of texture or detectable features. An implementation of SIFT (Scale Invariant Feature Transform) [18] features by VLFeat.org has been used for rectification to obtain the initial matches.

The SIFT detector extracts from an image a number of frames or attributed regions in a way which is consistent with some variations of the illumination, viewpoint and other viewing conditions. The descriptor associates to the regions a signature which identifies their appearance compactly and robustly. About seven thousand matching features per image pair had been produced by SIFT. Finally, epipolar constraints were applied on those initial matching points to derive valid matches. An example of the feature matches derived by using SIFT features and applying epipolar constraints is shown in Figure 3. Only forty matching points had been plotted on the image pair for illustration. More than six thousand points were successfully matched for this image pair, after starting with over seventeen thousand pairs, by repeatedly using the constraints to confirm the match. The final matches have been verified by human users.

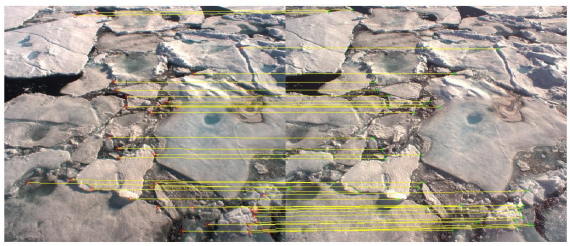


Figure 3: Matched features of stereo image pair by applying SIFT and epipolar constraints with only 40 matches shown

4. Methods

Fifteen image pairs were selected for this analysis. Block matching was applied using different metrics for detecting a valid match - Sum of Absolute Differences (SAD) and Normalized Cross Correlation (NCC). Block matching requires taking a small region of pixels in the left image, and searching for the closest matching region of pixels in the right image. Multiple experiments have been performed on various images and on various regions within the image pairs to reach at the conclusion that the maximum shift of features between the left and the right is no more than 32 pixels.

Varying this template size produces interesting results. Increasing the template size, produces noisy but accurate disparity maps, while decreasing creates less noisy however less accurate results. Templates of sizes varying from 5x5 to 45x45 have been tested using repeated experiments. The optimal template size for the ice dataset is between 19x19 and 23x23, which produced results that are not too noisy yet accurate. The rectified image pairs are converted to grey-scale before stereo analysis so that each pixel has only one value (0 - 255) instead of three (R,G,B). Two types of algorithms have been used and described below.

4.1. Local Algorithms

For local algorithms, the WTA (Winner-takes-All) approach for optimization had been used. To find the most similar block to the template, SAD values were computed between the template and each block in the search region, then the block with the lowest SAD value was chosen.

The Sum of Squared Differences (SSD) metric is more sensitive than SAD because it uses squared error terms instead of absolute values. So the resulting disparity maps were produced with more noise for SSD metric. For Normalized Cross Correlation (NCC), the mean is subtracted from the SSD and divided by the standard deviation. NCC provided better results than SSD in case of ice dataset. Block matching had been utilized in this work with SAD and NCC metrics.

4.2. Global Algorithms

Unlike the local algorithms, while using the global algorithms we skip the cost aggregation and define a global energy function. Dynamic programming processes the stereo matching using two dimensional optimization, e.g., Belief Propagation (BP) [35] and Graph Cut (GC) [36]. Blocks were compared using Semi Global Block Matching, instead of pixel-wise comparison. Matching cost was hierarchically calculated by Mutual Information (MI) [37, 38, 39]. MI was chosen because it is insensitive to recording and illumination changes and is defined from the entropy of two images, as well as their joint entropy. The matching cost calculation added a smoothness constraint by penalizing changes in neighboring disparities to prevent wrong block matches. A disparity image corresponding to base image and one to match image were then calculated. Outliers were filtered by a small 3 x 3 window. Consistency check was performed based on the basis of whether both disparities differ, which enforced the uniqueness constraint by permitting only one-to-one mappings.

4.3. Experiments

The dataset had been used to select a total of fifteen stereo image pairs for the experiment. After rectifying each image pair, it had been verified if features line up on the Y-coordinates. If not, the image pair was run through the rectification process again. Some pairs had to be discarded due to the failure of the rectification process even after multiple trials. Difference in light and shade, other than

occlusion could cause such problems.

Because ground truth was not available for the dataset, disparity value for each image set had been verified by clicking on the same feature points of left and right rectified images. These points were then used in lieu of the ground truth values to compare each algorithm's accuracy on each of the grouped regions as described below.

For each extracted point pair (left and right), kernels of sizes varying between 3x3 to 15x15 were created. Two different metrics - mean difference and normalized mutual information (NMI) were calculated and tested for each pair of kernels. Gaussian Mixture Model (GMM) was applied on these two sets of metrics derived from the image pair matching. GMM was used to test for the underlying distributions that these patches might have been generated from. Histograms of occurrences of these metrics were plotted separately and compared against each other to find evidence for the number of underlying distributions. GMM was chosen over k-means clustering due to the nature of the data and the ability of GMM to provide better results as shown by Priebe et al. [40]. From the various combinations of metrics and kernel sizes, it was found that the NMI metric for kernel size of 9x9 produced the most coherent normal shaped histogram as shown in Figure 5. The histograms obtained using other kernel sizes and different metric did not produce satisfactory results. Those histograms suffered from abnormal gaps or unusual shapes of the graphs. The most optimal number of distributions were checked starting from two to twelve. Histograms with three underlying distributions came up with the most satisfactory characteristics, which had the least overlap with other groups and almost no gap between groups. The NMI values range between [0,1] as the name suggests. The X-axis shows the NMI and the Y-axis plots the number of patches with that NMI value. Group one has 6,678, two has 1,694 and three has 7,625 data points in Figure 5 for a total of 15,997 data points. The number of data points are justified for the Central Limit Theorem to be in effect and get a reasonably good statistical model. Hence three regions per image are obtained using GMM on NMI scores of kernel pairs.

For block matching algorithms, different block sizes starting from 5x5 to 23x23 were tried to determine the best block size for ice dataset. For SAD metric 7x7 and for the rest 19x19 block sizes produced best results when comparing error rates for disparity. The growing seed method did not require block size to be supplied.

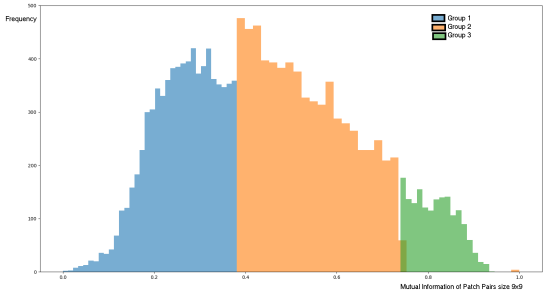


Figure 5: Histogram of Patch Similarity using Normalized Mutual Information

4.3.1 Using Local and Global Algorithms

The algorithms were applied to each stereo image pair after rectification. The disparity estimates were extracted for each of the four algorithms. Over fifteen thousand points had been clicked and verified on the image pairs for the experiment, which were uniformly spread across the entire image. The four stereo methods that produced best results on ice were: Block Matching with Sum of Absolute Differences (SAD), Block Matching with Normalized Cross Correlation (NCC), Semi-Global Block Matching (SGBM), and a seed growing method with global optimization (GCS). The error rates for these four algorithms were the lowest among all others tested for the purpose.

4.3.2 Using Algorithm Output

After employing the above method, errors were computed using the generated ground truth (GT) values. For each of the points that had GT, the Root Mean Squared Error (RMSE) was computed for each of the four methods by group. The RMSE provided an idea of how much error each algorithm made on each of the three grouped regions. Table 1 demonstrates the results, which were used in making a decision about which method performs better for each region that was grouped by GMM. It was hypothesized that different algorithms might perform differently on the regions, because those belong to different statistical distributions identified by the GMM model.

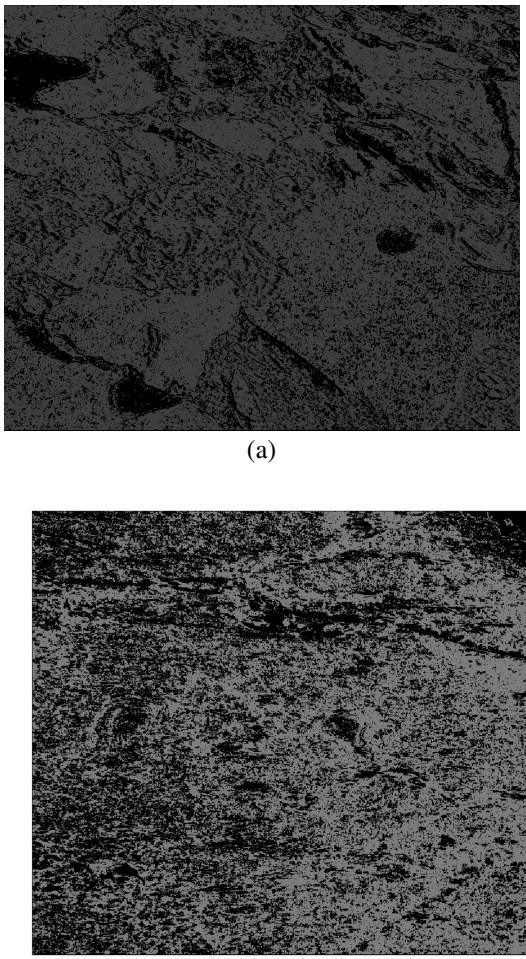
ALG/GROUP	One	Two	Three
GCS	15.47	10.21	2.23
SGBM	13.75	0.15	11.87
NCC	2.99	1.83	3.04
SAD	6.41	3.20	5.16

Table 1: RMSE (in pixels) for each algorithm by group

It could be inferred from the above table that NCC produced best results on group one (with 1,694 points), SGBM on two (with 6,678 points) and GCS on three (with 7,625 points). Hence, these algorithms were invoked for each of the groups for generating the most accurate disparity map for ice images.

5. Results

The output disparity map for two methods has been displayed in Figure 6.



(a)

(b)

Figure 6: Disparity Maps using - (a) SAD (b) GCS

When compared with the rectified image pair in Figure 3, it can be noticed that individually the algorithms work better in certain regions, but not overall.

Finally, three sets of image pairs were taken from the ice dataset. GT for 11,971 points were extracted by using SIFT features and epipolar constraints, and then verified and recorded. Using the stereo algorithms, the MI for 9x9 matching kernels were computed to determine which group these points belong to. This information was used to select the method that performs best on that group. The final shift in correspondence points was estimated using the selected algorithm to generate the values for those groups. RMSE was calculated for each algorithm by group and shown here in Figure 7. The distribution of RMSE over the groups confirmed the choice of algorithms selected by GMM that uses NMI metric to evaluate the kernel matches for the stereo pairs. NCC produced best results on region one, SGBM on region two and GCS on region three. Roughly 1000 square meter of scene is captured by the camera sensors and an error of 3 pixels in Figure 7 translates to approximately 3 centimeter, which is our largest RMS error while measuring the thickness of the ice layer.

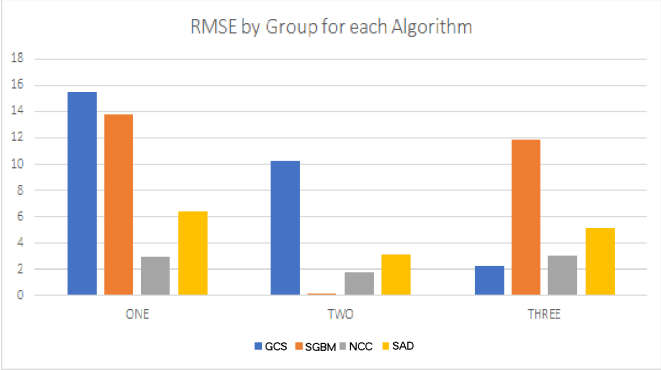


Figure 7: RMSE (Y-axis in pixels) for each Algorithm by Group

6. Conclusions

A novel approach has been developed for stereo analysis of ice images, which suffer from low texture resulting in poor matches and inability to rectify. A method has been proposed for grouping regions of the ice images using Gaussian Mixture Models to predict the probable distributions that the pixels may have been generated from. A suitable kernel size and metric are chosen, on which GMM is applied. The accuracy has been evaluated quantitatively and qualitatively.

Unavailability of ground truth poses a problem and has been addressed using methods to generate ground truth by sampling sufficient amount of points throughout the image pairs. Use of a LIDAR for the purpose has been considered however, the cost point will be very different for the

experiment. Currently two 5 MP FLIR Point Grey cameras have been used in a heated fireproof casing, mounted on the flying deck of the ship. Table 2 shows specifications and pricing data on LIDAR units and the camera that was used for PSITRES stereo images.

Device	HDL-64E	AlphaPuck	OS2	SX10	FLIR
Range(meter)	120	300	240	800	1000
No. of Lines	64	128	64	1000	2048
MM Pts/Sec	2.2	9.6	1.31	1.11	110
FPS	20	20	20	15	22
Price(USD)	75,000	8,000	24,000	38,765	835

Table 2: Four LIDAR and FLIR Camera Specifications

Use of LIDAR poses some challenges other than just price. With relatively low speed of acquisition for large distances, it can be challenging to use these units on a cruise ship that maintains high speeds throughout its course. One of the objectives of the project is to have fully configured systems mounted on such ships to collect data and bring those back for research purposes like stereo reconstruction. Currently the ice-going research vessels travel at varying speed and can stop for certain duration for image acquisition. The lower acquisition speed may result in erroneous data, although this might improve in near future. It might take more than one second to record ground truth for one scene with any of the LIDAR units mentioned in the table. In recent days, LIDAR units have been used on self-driving cars; however, those are primarily for collision avoidance. Collision happens only at very close range, which is not suitable for our research. The recording range is also on the lower side for LIDAR, as can be seen in the Table 2. The stereo cameras can both capture at high speed (due to higher frame rate) and objects that are farther than 1000 meter from the unit. The cruise ships are supposed to stay further away from the sea ice than the ice-going research vessels. This gives an advantage for using stereo camera units over LIDAR units.

In the future, we intend to acquire ground truth by using a LIDAR unit during a research expedition. This will be strictly used for building a model to train for performing stereo analysis and then test the model before deployment, where LIDAR units will not be present.

References

[1] Karen E Frey, Donald K Perovich, and Bonnie Light. The spatial distribution of solar radiation under a melting arctic sea ice cover. *Geophysical Research Letters*, 38(22), 2011.

[2] Michiyo Yamamoto-Kawai, Fiona A McLaughlin, Eddy C Carmack, Shigeto Nishino, and Koji Shimada. Aragonite undersaturation in the arctic ocean: effects of ocean acidification and sea ice melt. *Science*, 326(5956):1098–1100, 2009.

[3] Michael Steele, Jinlun Zhang, and Wendy Ermold. Mechanisms of summertime upper arctic ocean warming and the effect on sea ice melt. *Journal of Geophysical Research: Oceans*, 115(C11), 2010.

[4] HDx Pritchard, SRM Ligtenberg, HA Fricker, DG Vaughan, MR Van den Broeke, and L Padman. Antarctic ice-sheet loss driven by basal melting of ice shelves. *Nature*, 484(7395):502, 2012.

[5] Scott Sorensen. Computer vision for polar sciences.

[6] D. Scharstein; R. Szeliski. Middlebury stereo vision page.

[7] Kuk-Jin Yoon; In So Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 650 – 656, 2006.

[8] Rowel Atienza. Fast disparity estimation using dense networks. *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[9] Ang Li, Dapeng Chen, Yuanliu Liu, and Zejian Yuan. Coordinating multiple disparity proposals for stereo computation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4022–4030, 2016.

[10] Heiko Hirschmüller, Peter R. Innocent, and Jon Garibaldi. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, 47(1):229–246, Apr 2002.

[11] Christoph Strecha, Wolfgang Von Hansen, Luc Van Gool, Pascal Fua, and Ulrich Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. Ieee, 2008.

[12] Jure Zbontar and Yann LeCun. Computing the stereo matching cost with a convolutional neural network. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2015.

[13] Haesol Park and Kyoung Mu Lee. Look wider to match image patches with convolutional neural networks. *IEEE Signal Processing Letters*, 24(12):17881792, Dec 2017.

[14] Amit Shaked and Lior Wolf. Improved stereo matching with constant highway networks and reflective confidence learning, 2016.

[15] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks, 2015.

[16] Xiaoqing Ye, Jiamao Li, Han Wang, Hexiao Huang, and Xiaolin Zhang. Efficient stereo matching leveraging deep local and context information. *IEEE Access*, PP:1–1, 09 2017.

[17] Jure Zbontar and Yann LeCun. Computing the stereo matching cost with a convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1592–1599, 2015.

[18] M.E. Munich ; P. Pirjanian ; E. Di Bernardo ; L. Goncalves ; N. Karlsson ; D. Lowe. Sift-ing through features with vipr. *IEEE Robotics Automation Magazine*, 2006.

[19] Jan Čech and Radim Šára. Efficient sampling of disparity space for fast and accurate matching. In *BenCOS 2007: CVPR Workshop Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images*. IEEE, 2007.

[20] Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. End-to-end learning of geometry and context for deep stereo regression, 2017.

[21] Jiahao Pang, Wenxiu Sun, Jimmy SJ. Ren, Chengxi Yang, and Qiong Yan. Cascade residual learning: A two-stage convolutional neural network for stereo matching, 2017.

[22] Patrick Knbelreiter, Christian Reinbacher, Alexander Shekhovtsov, and Thomas Pock. End-to-end training of hybrid cnn-crf models for stereo, 2016.

[23] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

[24] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):328–341, February 2008.

[25] Daniel; Szeliski Richard Sinha, Sudipta; Scharstein. Efficient high-resolution stereo matching using local plane sweeps. *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[26] C Kim; H Zimmer; Y Pritch; A Sorkine-Hornung; Markus Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM ToG*, 2013.

[27] Konstantinos Batsos, Changjiang Cai, and Philipos Mordohai. Cbmv: A coalesced bidirectional matching volume for disparity estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[28] Aristotle Spyropoulos and Philippos Mordohai. Correctness prediction, accuracy improvement and generalization of stereo matching using supervised learning. *International Journal of Computer Vision*, 118, 12 2015.

[29] Aristotle Spyropoulos and Philippos Mordohai. Ensemble classifier for combining stereo matching algorithms. In *2015 International Conference on 3D Vision*, pages 73–81. IEEE, 2015.

[30] Matteo Poggi and Stefano Mattoccia. Learning a general-purpose confidence measure based on o (1) features and a smarter aggregation strategy for semi global matching. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 509–518. IEEE, 2016.

[31] Akihito Seki and Marc Pollefeys. Patch based confidence prediction for dense disparity map. In *BMVC*, volume 2, page 4, 2016.

[32] Matteo Poggi and Stefano Mattoccia. Deep stereo fusion: combining multiple disparity hypotheses with deep-learning. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 138–147. IEEE, 2016.

[33] Min-Gyu Park and Kuk-Jin Yoon. Leveraging stereo matching with learning-based confidence measures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 101–109, 2015.

[34] Spyros Gidaris and Nikos Komodakis. Detect, replace, refine: Deep structured prediction for pixel wise labeling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5248–5257, 2017.

[35] Alexander T. Ihler, John W. Fischer III, and Alan S. Willsky. Loopy belief propagation: Convergence and effects of message errors. *J. Mach. Learn. Res.*, 6:905–936, December 2005.

[36] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, November 2001.

[37] Stephen J Krotosky and Mohan M Trivedi. Mutual information based registration of multimodal stereo videos for person tracking. *Computer Vision and Image Understanding*, 106(2-3):270–287, 2007.

[38] Junhwan Kim, Vladimir Kolmogorov, and Ramin Zabih. Visual correspondence using energy minimization and mutual information. In *null*, page 1033. IEEE, 2003.

[39] Clinton Fookes, A Maeder, Sridha Sridharan, and Jamie Cook. Multi-spectral stereo image matching using mutual information. In *Proceedings. 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004.*, pages 961–968. IEEE, 2004.

[40] Carey E. Priebe, Youngser Park, Joshua T. Vogelstein, John M. Conroy, Vince Lyzinski, Minh Tang, Avanti Athreya, Joshua Cape, and Eric Bridgeford. On a two-truths phenomenon in spectral graph clustering. *Proceedings of the National Academy of Sciences*, 116(13):5995–6000, 2019.