Robotics and Autonomous Systems 87 (2017) 247–257

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Multi-Channel Generalized-ICP: A robust framework for multi-channel scan registration

James Servos, Steven L. Waslander*

Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON, Canada, N2L 3G1

HIGHLIGHTS

- Scan registration method that employs additional channels of information like color, intensity.
- Integrates new channels into point covariance weights used in registration optimization.
- Improves registration in regions with limited geometric features, without computation penalty.
- Demonstrated on Ford, Freiburg, and Waterloo hallway datasets.

ARTICLE INFO

Article history:
Received 30 June 2015
Received in revised form 12 June 2016
Accepted 20 October 2016
Available online 26 October 2016

Keywords:
Scan registration
RGBD point clouds

ABSTRACT

Current state of the art scan registration algorithms which use only position information often fall victim to correspondence ambiguity and degeneracy in the optimization solutions. Other methods which use additional channels, such as color or intensity, often use only a small fraction of the available information and ignore the underlying structural information of the added channels. The proposed method incorporates the additional channels directly into the scan registration formulation to provide information within the plane of the surface. This is achieved by calculating the uncertainty both along and perpendicular to the local surface at each point and calculating nearest neighbor correspondences in the higher dimensional space. The proposed method reduces instances of degenerate transformation estimates and improves both registration accuracy and convergence rate. The method is tested on the Ford Vision and Lidar dataset using both color and intensity channels, as well as with Microsoft Kinect data from the Freiburg RGBD Office dataset and data obtained from the University of Waterloo campus.

1. Introduction

Simultaneous localization and mapping (SLAM) is one of the cornerstone components of autonomous systems operating in an unknown environment. With many modern sensors, such as LIDAR, RGBD cameras, and stereo cameras, providing robots with reliable 3D point information of their environment, scan registration techniques have become a prevalent front end solution to the SLAM problem. Scan registration aligns pairs of scans to obtain the rotation and translation of the system relative to its environment between the two poses. These measurements can then be combined in a pose graph optimization back end, which refines the SLAM solution and allows for the aggregation of point cloud data into aggregated maps. These aggregate maps provide detailed environmental information which can be used for path planning and obstacle avoidance.

Classic scan registration formulations such as the Iterative Closest Point (ICP) [1], Generalized ICP (GICP) [2] and Normal Distribution Transform (NDT) [3] algorithms use only 3D point information to calculate point correspondences, distributions, and to perform the registration. These methods have been demonstrated to work well in a wide range of environments, but registration accuracy suffers when the environment is devoid of rich geometric variation. However, many sensors, or combinations of sensors, can provide additional information for each point such as intensity or color. The inclusion of the additional information channels can help to improve registration accuracy, convergence rate, and solve many structural ambiguities such as occur in hallways or repetitive environments.

Previous work has been performed to include additional channels into the scan registration method. Intensity information is often available from LIDAR system and color information has become commonly applied to 3D scans. Cameras and depth sensors can be extrinsically calibrated such that points can be associated with corresponding image pixels and colorized. A multitude of

* Corresponding author.
E-mail addresses: servos@gmail.com (J. Servos), stevenw@uwaterloo.ca (S.L. Waslander).

http://dx.doi.org/10.1016/j.robot.2016.10.016

variations of the ICP method have been proposed over the years, many of which are compared in [4], and several have attempted to incorporate additional information. Most of these methods use the original ICP cost function and incorporate the information by adding additional constraints or dimensions to the correspondence computation including [5,6] and [7] which incorporate color information as well as surface descriptors, and, [8] and [9] which focus on intensity information. Color information has also been proposed to be incorporated into NDT in Color NDT [10] which has been shown to improve upon the geometry-only methods but still have several drawbacks. One of the major issues with incorporating secondary information, particularly color, is the inclusion of both noise and illumination changes.

The Multi-Channel Generalized-ICP (MCGICP) method, first proposed in [11], is an extension of the GICP method which incorporates additional channels of information for each point. The additional channels are used to generate a spatial covariance in the plane of the surface which is used to compliment the existing plane to plane matching and allow the planes to be aligned not only in the normal direction, but also perpendicular to it. The in-plane covariance is calculated using a kernel weighted covariance based on the additional channels and is normalized by the unweighted population covariance of the points. The in-plane covariance is added to the GICP archetypal covariance in the planar directions and is then rotated into 3D space as performed in GICP. Point correspondences are also changed to use the Color ICP method, which leverages a higher dimensional weighted kD tree. The error function is unchanged from the original GICP, however the modified covariance will induce non-trivial error terms in the planar directions.

In this work we expand upon our original work to improve the formulation of the MCGICP method as well as provide additional experimental validation of the results. First, we provide an improved and more thorough explanation of the method as well as the theoretical decomposition of the algorithm in common extrema cases in order to demonstrate its stability analytically. Second, we propose the use of eigenvalues as an additional term in the correspondence calculations in order to improve correspondence robustness and accuracy. Finally, an in-depth analysis of the results of the method is presented on multiple datasets including, the Ford Campus Vision and Lidar dataset [12], the Freiburg RGBD Office dataset [13], and a dataset collected at the University of Waterloo campus of Microsoft Kinect data in a hallway environment.

2. Related work

Scan registration methods for 3D point clouds are well established. One of the first methods proposed for solving the scan registration problem was the Iterative Closest Point (ICP) method introduced by Besl and McKay [1]. The ICP algorithm minimizes the Euclidean distance between nearest neighbor points in the two scans to find the relative transform. Taking advantage of the locally planar nature of most environments, Chen and Medioni [14], proposed a point to plane based variant of ICP which penalizes the cost only normal to the surface of the environment. This approach mitigates the sampling error seen in point to point ICP which assumes that points correspond exactly between scans. More recently, Segal et al. developed Generalized-ICP (GICP) [2] which, using a probabilistic framework, generalizes the ICP method and introduces a plane to plane approach with improved performance over the previous versions. An alternative approach, called the normal distribution transform (NDT), was first suggested by Biber and Strasser in [3] in 2D and extended to 3D by Magnusson et al. [15]. NDT segments the scan into fixed size voxels and calculates a normal distribution of the points within each cell. Scans are then registered by using the point to distribution or distribution to distribution [16] error metric.

Although effective in many environments, the purely point based registration methods miss the opportunity of using additional sources of information to disambiguate degenerate environments. Color ICP [5], and similar related works [6,8,9], attempt to improve ICP performance by using the additional channels, such as color, to perform the nearest neighbor search in a higher dimension or by constraining the search space [7]. This method showed improved point correspondence results but did not change the underlying scan registration method. A colorized version of NDT has also been proposed by Huhle et al. [10], which uses color based kernel functions to generate a Gaussian mixture model such that each voxel then contains a color based mixture of Gaussians. Due to the voxelization of the environment this method can lose some of the finer details of the environment structure. This can cause reduced accuracy or even registration failure when these details are vital to determining a correct registration. Servos and Waslander proposed an extension to the Color NDT algorithm, Color Clustered NDT (CCNDT), in [17] which uses color clustering to generate a set of normal distributions instead of the standard voxel grid. This method improves computational performance and accuracy but requires an environment with sufficient color variation for clustering.

An alternate approach to incorporate both color and depth information in point cloud registration uses image features augmented with depth information. Inspired by low-cost RGBD sensors and stereo vision data, both color and depth information is available at every pixel in the image. RGBD-SLAM [18], has been developed by Endres et al. and uses RGBD images collected from a Kinect or stereo camera to track SIFT or SURF image features in 3D space and optimize the motion transform using a graph SLAM back-end. This approach can produce accurate maps but can slow down significantly as the graph grows and can struggle is non-feature-rich environments. Stereo vision has also been successfully used for real-time SLAM by Konolige et al. [19], who augmented stereo feature matching with sparse bundle adjustment, which provides accurate pose information in rough outdoor environments over large trajectory lengths with relatively low position drift. Other methods combine scan registration and image feature correspondence to improve results. These methods include, [20], which initialize GICP using image features and [21] which augments NDT using a small set of image feature correspondences as a secondary error function.

Recently, with the development of more powerful computers and parallel processing environments such as CUDA [22], it has become feasible to perform dense SLAM using all the information captured in an image frame. These methods attempt to use every pixel in an image to estimate vehicle pose. Therefore they can be extremely computationally expensive, however, with parallelization are often capable of running in real time.

Several dense camera SLAM methods have been proposed including Dense Tracking and Mapping (DTAM) [23], created by Newcombe et al., which uses a single monocular camera to generate 3D reconstructions of environments. This method creates accurate reconstructions but cannot determine scale without an outside source and is ideal for close up applications. Newcombe et al. also proposed the dense SLAM method KinectFusion [24], which uses a Kinect sensor to generate dense 3D maps of small areas. The method was further extended by Whelan et al. [25] to allow larger working volumes and include color. These methods rely on GPU parallelization in order to be able to perform in real time, and can be extremely memory intensive as the mapped region grows.

While image based techniques have the advantage of providing long distance bearing measurements, they suffer from poor range and field of view when compared to the laser based approaches. Camera based methods typically require detailed close quarters information in order to perform effectively and struggle to handle

larger, more homogeneous environments. They generally have significant difficulty in making large quick movements, due to motion blur and tracking, and using long range measurements, due to poor triangulation, such as are expected in outdoor operations.

In general, feature based methods can have several draw backs as demonstrated in [10]. Since feature based methods rely on matching a small number of points, noise in the 3D location can cause significant errors. Additionally, a small number or even a single false correspondence can cause catastrophic failure without significant reliance on robustification methods in the SLAM back end. This can make feature based methods unreliable particularly in dynamic environments where a single, correctly corresponded feature could move and distort the entire map. Scan registration methods on the other hand can fail due to a lack of geometric texture which can cause a degeneracy in the optimization and rely heavily on the structure of the environment in order to perform optimally.

The addition of color information into the scan registration formulation is clearly of benefit and is able to mitigate some of the major issues present in the classic scan registration algorithms as well as improve accuracy and run time in some cases. The work presented in this paper builds upon these colorized scan registration approaches such as Color ICP, and generalizes the methods into a robust framework for reliable scan registration for any multi-channel sensor suite being used for robot autonomy.

3. Problem formulation

Scan registration algorithms attempt to find the optimal transform between an input scan and a reference scan. The optimal transform is qualitatively defined as the transform which best aligns the scans into a common coordinate frame. Given scan $A = \{a_i\}$, where $a_i \in \mathbb{R}^3$ for $i = 1, \ldots, N_A$, and scan $B = \{b_j\}$, where $b_j \in \mathbb{R}^3$ for $j = 1, \ldots, N_B$, the optimal transformation matrix, T, can be defined as,

$$T = \begin{bmatrix} R_T & t_T \\ 0 & 1 \end{bmatrix}$$

where $R_T \in SO(3)$ is the rotation matrix and $t_T \in \mathbb{R}^3$ is the translation vector.

The solution is typically computed by optimizing a specified score function, $\Gamma : SE(3) \rightarrow \mathbb{R}$, given the point clouds A and B, to find the optimal transformation T. The optimization can be generally defined in the form:

$$T^* = \arg\min_T \Gamma(T). \tag{1}$$

The score function is specific to each algorithm and can take many different forms and in many cases is optimized in an iterative method before settling on the final solution. The following section will introduce the score function definition for GICP, used as a basis for the MCGICP algorithm. For other commonly used scan registration methods see [1] for standard ICP, and [3] for NDT.

4. GICP

The Generalized Iterative Closest Point (GICP) method was developed by Segal et al. in [2] as a unifying framework of the previously proposed ICP methods. The GICP formulation uses a probabilistic framework to determine the error function and proceeds as follows.

First, it is assumed that the nearest neighbor correspondences have been calculated and scan A and scan B are indexed with corresponding points having the same indices and non corresponding points being removed. Using the probabilistic model it is assumed that the point clouds A and B are generated from an underlying set of distributions, where $a_i \sim \mathcal{N}(\hat{a}_i, C_i^A)$ and $b_i \sim \mathcal{N}(\hat{b}_i, C_i^B)$. Therefore, given perfect correspondences and the correct transform, T^*,

$$\hat{b}_i = T^* \hat{a}_i. \tag{2}$$

The difference between samples a_i and b_i is then defined as $d_i = b_i - Ta_i$. Given that a_i and b_i are drawn from independent Gaussian distributions, and given the correct transformation, d_i can be written as:

$$d_i \sim \mathcal{N}(0, C_i^B + T^* C_i^A (T^*)^T). \tag{3}$$

The transform is then solved for using maximum likelihood estimation (MLE) and simplified to the form

$$T = \arg\min_T \sum_i d_i (C_i^B + T C_i^A T^T) d_i^T. \tag{4}$$

This formulation can be used to represent any of the standard forms of ICP including basic point to point as well as point to plane ICP. However GICP proposes a plane to plane model in which it is assumed that points are sampled from surfaces which are locally planar. In this model the covariance of a point is assumed to be small in the direction of the normal at that point and large in all other directions. This assumes that the points have little information to offer in the directions tangent to the plane. The covariance at every point, $q_i \in A \cup B$, in both A and B is calculated using an archetype covariance, C^G, define as

$$C^G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \epsilon \end{bmatrix}$$

where ϵ is a constant representing the covariance along the normal. The covariance at a point is then calculated as

$$C_i^W = (R_i^W) C^G (R_i^W)^T \tag{5}$$

where R^W is the rotation matrix which rotates ϵ to align with the surface normal, at point q_i.

The local covariance, C_i^L, is calculated using the k nearest points to the query point q_i found using [26]. The local covariance approximates the model covariance in the region around the query point. The surface normal information for this method is then computed using principal component analysis (PCA) on the local covariance, C_i^L. The component with the lowest eigenvalue corresponds to the surface normal. In practice, the model covariance at a given point can be calculated using the singular value decomposition (SVD) of the local covariance,

$$C^L = USV^T \tag{6}$$

where the singular values are the diagonal elements of $S \in \mathbb{R}^{n \times n}$ sorted in descending order and U and V are orthonormal matrices. In the singular value decomposition, U is equivalent to the rotation matrix R_i^W and therefore S can be replaced by C^G, to compute, C_i^W.

Fig. 1 shows an example of the alignment of two planar scans using the GICP algorithm. The plane to plane nature of GICP allows the scans to be aligned correctly without falling into a local minima, which tends to happen more frequently in the case of standard ICP where local normals are not included.

5. Multi-channel GICP

The proposed method, MCGICP, is an extension of the GICP algorithm which incorporates additional channels of information. The MCGICP algorithm uses additional channels such as color, intensity, or any other spectral information, to introduce additional information to the problem. Additionally, MCGICP uses the added channels directly in the correspondence search to attempt to provide more robust results. The increased problem space not only solves the degeneracy problem but also improves accuracy, convergence, and robustness of the scan registration results without significantly increasing computational complexity.

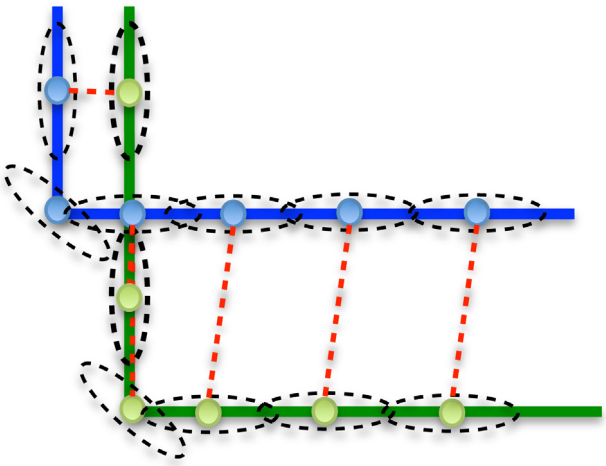


Fig. 1. Example of the alignment of two scans (green and blue) using the GICP algorithm. As can be seen each point has a corresponding covariance which aligns with the surface of the scan (black). Corresponding points (red lines) with covariances which are aligned will result in a higher weighted cost in the minimal direction compared to those which are not aligned. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

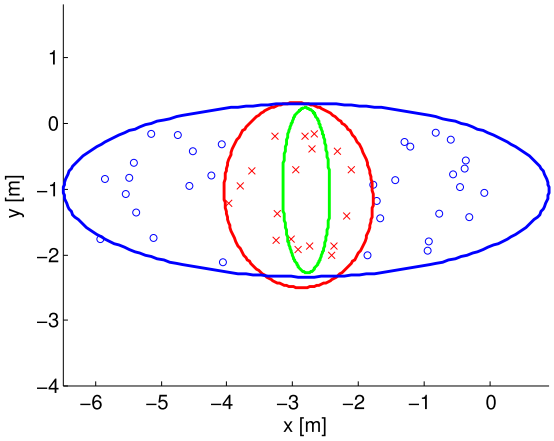


Fig. 2. An example of the population covariance (blue), and descriptor covariance for the biased distribution (red), and the resulting normalized correlation, Ω (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.1. Covariance calculation

MCGICP assumes, as GICP did previously, that the environment is locally planar and that the 3D points only contain useful information in the direction normal to the surface. However, since the points have at least one additional channel of information, the additional channel(s) can be used to define the covariance of a point along the surface plane as well. The added channels will have no effect normal to the plane as the sample must lie on the surface and therefore will complement the positional information well.

First, let all points, $p_i = \{p_i^p, p_i^d\}^T$, have both positional information, $p_i^p \in \mathbb{R}^3$, and n descriptor channels, $p_i^d \in \mathbb{R}^n$, which can include, for example, intensity or RGB color information. Then for each point in point clouds A and B, the model covariance sets, C^A and C^B, are calculated using both position and descriptor information.

Let $q \in A \cup B$ be the current query point for which the model covariance is to be calculated. The local covariance, C_i^L, of the query point position is calculated using the k nearest neighbor points to q^p using the fast local approximate nearest neighbor algorithm [26]. Let the nearest neighbors be defined as the set of points $L = \{l_j\}$ for $j \in \{1, \ldots, k\}$, such that $\|l_j^p - q^p\| \leq \|r - q^p\|$, for all $r \in Q \cap \bar{L}$, where Q is the point cloud associated with the current query point, q. Singular value decomposition (SVD) is performed to extract the principal components. The normal of the surface is found as the component with the smallest singular value in S. The neighborhood points are then projected onto the plane perpendicular to the normal and reduced to $\mathbb{R}^2$. Let $z_j^p \in \mathbb{R}^2$ be the projected point and $Z = \{z_j^p, z_j^d\}^T$ be the new set of points in $\mathbb{R}^{2+n}$. The projection is then given as

$$\begin{aligned} z_j^p &= \begin{bmatrix} U_1{}^T \\ U_2{}^T \end{bmatrix} l_j^p \\ z_j^d &= l_j^d \end{aligned} \tag{7}$$

where U_1 and U_2 are the first and second columns of the SVD matrix U. Note that after the transform, the new population covariance, $\Sigma_w \in \mathbb{R}^{2\times 2}$, of the points, Z^p, is the diagonal matrix of the largest two singular values of S.

Now that the 3D points have been projected onto the local surface approximation, the reduction in uncertainty due to the incorporation of descriptor information in the plane can be calculated. To this end a descriptor kernel weighted covariance is calculated using weightings based on a probabilistic model similar to that used in [10]. The descriptor kernel calculates the probability that an arbitrary point corresponds to the query point in descriptor space. The kernel can be defined as a Gaussian distribution, $\mathcal{N}(q^d, \Lambda)$, centered at the query point descriptor, q^d, and with $\Lambda \in \mathbb{R}^{n\times n}$ being the measurement covariance of the descriptor sensor. The kernel weights are then calculated for each point in Z as:

$$\lambda_j = \exp\left(-\frac{1}{2}(z_j^t - q^t)^T \Lambda^{-1}(z_j^t - q^t)\right). \tag{8}$$

Using the kernel weights the descriptor kernel weighted covariance and mean, Σ_d and μ^p, can be calculated as

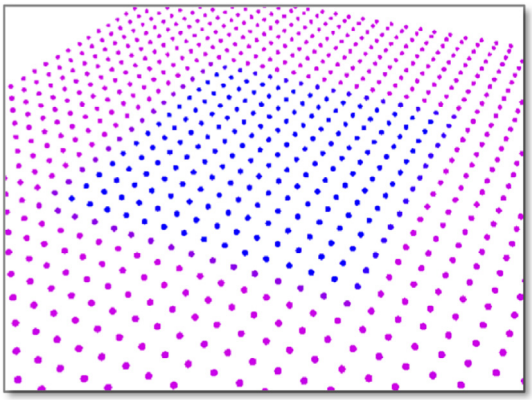
$$\mu^p = \frac{1}{\sum_j \lambda} \sum_j \lambda_j z_j^p \tag{9}$$

$$\Sigma_d = \frac{1}{\sum_j \lambda} \sum_j \lambda_j (z_j^p - \mu^p)(z_j^p - \mu^p)^T. \tag{10}$$

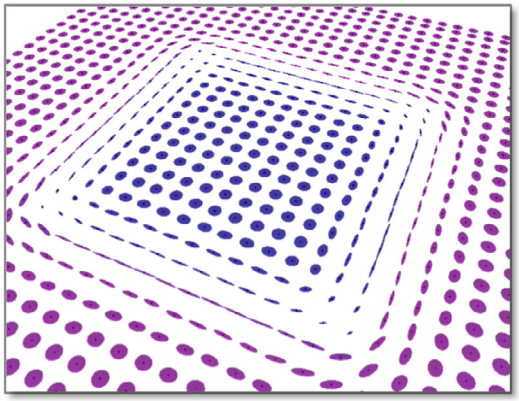
Eqs. (9) (10) give the spatial distribution of points based on their similarity to the query point. This distribution models the uncertainty of the descriptor information along the wall locally, however it can be biased if the original sample population was itself already biased. Fig. 2 shows an example of a population and descriptor covariance with a biased initial population. To compensate for this potential bias the distribution is normalized by the population covariance such that

$$\Omega = \Sigma_w^{-\frac{1}{2}} \Sigma_d \Sigma_w^{-\frac{1}{2}}. \tag{11}$$

The normalized descriptor covariance, $\Omega \in \mathbb{R}^{2\times 2}$, shows the correlation of the descriptor weighted data compared to that of the population. The normalization in this case is a Mahalanobis whitening transform on the population covariance. We use a pre and post multiplication of the square root of the inverse in order to ensure the matrix maintains symmetry. The population covariance would be transformed into an identity matrix in the normalized space and the descriptor covariance would be transformed relative to the population to give an unbiased descriptor covariance as if the population were identity. In practice a value less than one indicates that the descriptor data increased the data certainty in that direction, while a value less than one indicates an increase in uncertainty. Directions which have a low normalized covariance



(a) Example point cloud.



(b) MCGICP covariances.

Fig. 3. Example point cloud (left) with each point covariance calculated using the MCGICP method (right).

are more likely to capture correct correspondences in the descriptor space in that direction. Cases which have normalized covariances equal to or great than one indicate areas of low descriptor correspondence certainty, such as a wall of a single continuous color.

To use this information in the GICP framework, Ω is used along the planar directions. Therefore, the resulting covariance used in the MCGICP algorithms is

$$C^D = \begin{bmatrix} \Omega & 0 \\ 0 & \epsilon \end{bmatrix}.$$

Fig. 3 shows an example color point cloud surface with a distinctive color feature in the center and the corresponding covariances calculated by the MCGICP algorithm. It can be seen in the figure that points along the edges of the color boundary have covariances that align with the color edge. This allows the method to align the surface, even without geometric features, by aligning the edge features in the color channels.

The error function and minimization of the MCGICP method remain unchanged from the original GICP, presented in Eq. (4), which means that all current methods for solving the GICP optimization are still valid and no changes are necessary to the optimization.

5.2. Correspondence calculation

In addition to the covariance changes, the calculation of corresponding points is also changed to reflect the higher dimensionality of the information. An $n + 3$ dimensional weighted kD tree is used to incorporate all of the information into the search as first shown by Johnson and Kang [5]. A weighting vector, $\alpha = \{\alpha_1, \ldots, \alpha_{n+3}\}$ where $\alpha_1 = \alpha_2 = \alpha_3 = 1$ are the weights of the position data, is used to weight the descriptor information relative to the positional information.

The correspondence information does not need to come directly from the raw sensor information. Processed information can be used to improve the correspondences. By using the eigenvalues of the covariance matrix calculated at each point as well as spatial and descriptor information, better correspondence matching can be achieved. Using the eigenvalues of the covariance matrix weights the correspondence of points such that points with similarly shaped covariances will be corresponded. This results in points along the edges of both geometric and descriptor structures to be corresponded and can improve overall registration results. The use of other processed information to improve the correspondences further is possible, however this is left as a point of future work.

6. Extrema cases

Unlike many other methods, the MCGICP method is capable of gracefully handling the extreme cases of input scans. This includes the two practical instances which are a scan with a consistent descriptor value for every point and a scan with unique descriptor values for every point. Many other methods fail, and do not converge to the correct solution in one or more of these cases, where as MCGICP falls back to a logical state that is still capable of performing the registration with the information provided.

In the first case, where all points have identical descriptor information, the MCGICP method will mathematically become equivalent to the base GICP algorithm. This happens because the normalized descriptor covariance will become the identity when there is no change in the weighted descriptor covariance, Σ_d, compared to the population covariance, Σ_w. This would mean that,

$$\Sigma_w = \Sigma_d \tag{12}$$

$$\Omega = \Sigma_w^{-\frac{1}{2}} \Sigma_d \Sigma_w^{-\frac{1}{2}} = I \tag{13}$$

$$C^D = C^G. \tag{14}$$

Therefore the GICP archetype covariance, C^G, is assigned at every point. Since each point has the same descriptor and covariance, the correspondence search will only be dependent on the geometric distances. Thus the algorithm will be equivalent to the GICP algorithm. In this case, algorithms which rely extensively on unique descriptor information, such as feature based methods, will fail to register scans correctly.

In the second case, where all points have a unique descriptor, the MCGICP method will become equivalent to the color ICP algorithm. In this case, because all of the descriptors are unique, the weighted descriptor covariance will be very small. Thus it follows that,

$$\Omega \approx \epsilon I \tag{15}$$

$$C^D \approx \epsilon I. \tag{16}$$

Since C^D in this case is simply a scaled identity matrix and all points will have approximately the same covariance matrix, each point will have the same weighting relative to other points as well as the same weighting in all directions. The correspondence calculation will take into account these uniform descriptors with the result being that each point is corresponded by both descriptor and geometry. This is equivalent to how the color ICP algorithm would perform and in this case is a good approach for this type of environment. In this case algorithms such as GICP would fail to converge due to the lack of geometric features.

In both extreme cases the MCGICP algorithm seamlessly handles using the information available in an intelligent manner, allowing the method to be robust to the widest range of possible environments.

7. Implementation

The proposed method is a generalized framework which can be used with a variety of different sensor combinations. The number of additional channels which can be added to the points is not limited by the algorithm but in practice is only limited by diminishing returns on the usefulness of the information. Three possible sensor configurations are discussed below. First, a system which uses only a LIDAR sensor with intensity information, second, a typical colorized point cloud generated from a camera and range sensor combination, and finally a configuration incorporating a LIDAR sensor with intensity information with a camera setup to provide four additional information channels.

7.1. Laser intensity descriptor

A single LIDAR, such as the Velodyne HDL-64E, can add an additional channel in the form of laser intensity, which provides useful information to the registration. In the single channel case, the descriptor covariance reduces to a single value, and is found by computing the variance, $\Lambda \in \mathbb{R}$, of the laser scan intensity values. The weighting, α_4, of the intensity channel in the nearest neighbor search is used to scale the influence the intensity channel will have on finding nearest neighbors.

7.2. Color descriptor

The combination of a camera and a range sensor is a common setup on many robotic systems. Color provides three channels to incorporate into the model. In this case, $\Lambda \in \mathbb{R}^{3\times3}$ is a covariance matrix of three variables. However, it can be assumed that the color channels are independent and therefore Λ will be a diagonal matrix consisting only of the intra-channel variances. The exact values of Λ will depend not only on the sensor used but also on the color space which is chosen.

In this work, the RGB color space is used. In the RGB space, the color variances and weightings can be set equal to each other to represent equal uncertainty in each of the colors. The option of alternative color spaces is also possible as each space provides different benefits and considerations. For an HSV or YUV space, the variance of the value and illumination channels can be larger and the nearest neighbor weighting smaller to distinguish a higher uncertainty in illumination which is common in real world scenes. The choice of color space does not have a direct impact on the algorithm but does change the desired values of Λ and $\{\alpha_4, \alpha_5, \alpha_6\}$ and may impact performance depending on the characteristics of the environment and lighting. Recent work on lighting invariance for robotic perception [27] may also prove beneficial, but remain an area of future work.

7.3. Combined color and intensity

The combination of both color and laser intensity information presents an interesting configuration which is not typically leveraged in current algorithms. Although the channels of the combined descriptor could be considered to be independent it has been shown in [28] that laser intensity and color intensity are in fact positively correlated. This can be incorporated into the algorithm by setting the inter-channel covariance terms of Λ to non zero values. The covariance matrix can be determined experimentally using a set of known training data. The weighting values for the nearest neighbor search are typically inversely proportional to the variance of that particular channel and are therefore dependent on the specific sensors being used.

8. Experimental results

The MCGICP method is evaluated using three sets of data. The first set is the Ford Campus Vision and Lidar dataset [12]. The Ford dataset contains LIDAR and omni-directional image data as well as ground truth and is used to evaluate the quantitative accuracy of the scan registration results in outdoor environments. The method is evaluated using the laser intensity (I), color (C) and combined (C+I) descriptor channels as described in Section 7. The second dataset used for evaluation is the Freiburg RGBD dataset. The Freiburg dataset is generated in staged indoor environments and is collected using an RGBD camera. The Freiburg data is used to evaluate the accuracy and robustness of the method in indoor environments. The final set of data was obtained using a Microsoft Kinect sensor on the University of Waterloo campus. This dataset is used to evaluate the method on data with limited geometric structure. The method is evaluated qualitatively based on the reconstruction of a flat textured surface. Finally the convergence rate of each algorithm is compared using the Ford Campus dataset. In all cases the method is compared to both the original GICP and Color ICP algorithms as implemented in the Point Cloud Library (PCL) [29] as well as the CCNDT algorithm. In all evaluations the RGB color space is used.

8.1. Ford dataset absolute error

The Ford Campus Vision and Lidar dataset was generated using a Ford F-250 pickup truck equipped with a Velodyne HDL-64E laser scanner, a Point Grey Ladybug3 omni-directional camera, and an Applanix POS-LV 420 INS with Trimble GPS used for ground truth data. A series of 200 frames from a challenging subset of the dataset were used to perform pairwise registration on each consecutive pair of scans. This section was chosen as it contains limited geometry structure and represents a challenging area where previous methods can fail to converge due to degeneracy in the geometric constraints. No pose prior was used, so the initial estimate for the optimization was set to zero in all cases.

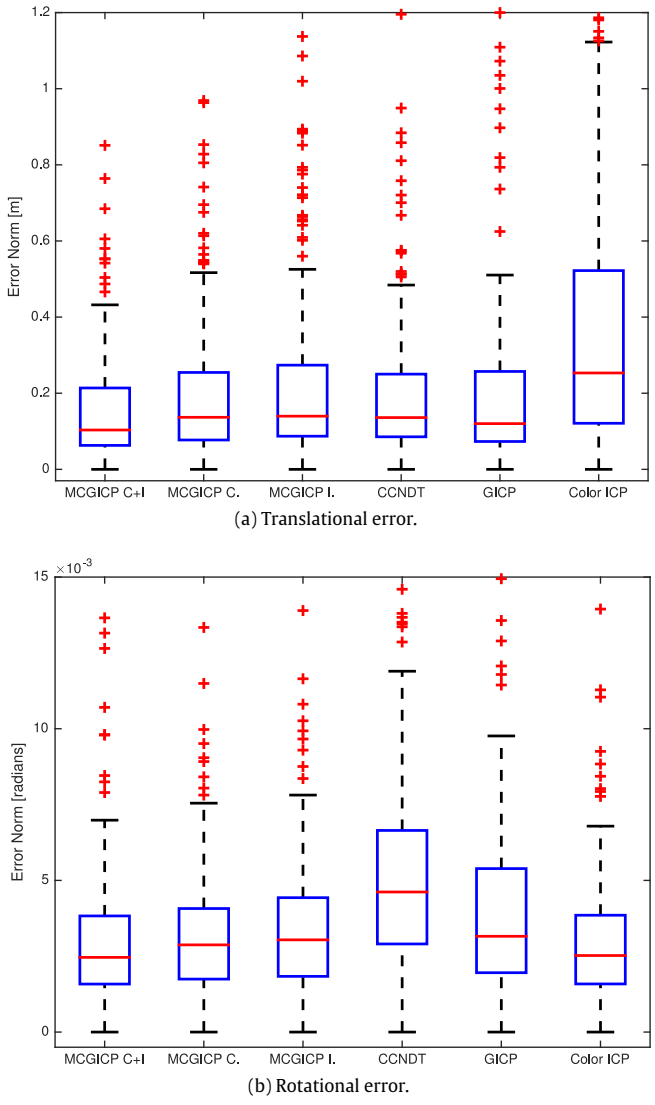
The ideal parameters used for each method were determined experimentally in order to produce the best results for this dataset. All methods used the BFGS optimization backend provided by PCL with termination conditions set to transformation epsilon 1e-6, euclidean fitness epsilon 1e-6 and max iterations of 50. GICP used a max correspondence distance of 1 m and Color ICP and MCGICP used a correspondence distance of 20 in the multi-channel descriptor space. CCNDT uses a color clustering threshold of 5 and a distance threshold of 2.0 with a minimum cluster size of 3. MCGICP used a weighting vector for the nearest neighbor search of $\alpha = \{\alpha_x, \alpha_y, \alpha_z, \alpha_R, \alpha_G, \alpha_B, \alpha_I, \alpha_e\} = \{1.0, 1.0, 1.0, 0.1, 0.1, 0.1, 0.05, 1.0\}$ and Color ICP used the same weighting excluding the intensity and eigenvalue terms. Finally, MCGICP used an idealized measurement model such that the diagonal of Λ was set to $\{120, 120, 120, 200\}$ corresponding to the RGB and intensity channels respectively, with zero cross correlation. An exact measurement model of the sensor could be generated by experimentation but it is expected that is would not improve results significantly and is left for future work.

The resulting transforms were then compared to the ground truth data to calculate the mean and standard deviation of the translational and rotational errors. The results are summarized in Table 1.

The error distributions from Fig. 4 as well as the error summary provided in Table 1 show that MCGICP using either color or combined descriptors has increased accuracy and reduced uncertainty over that of CCNDT, GICP, and Color ICP. Note that additional outliers exist outside the bounds of Fig. 4 but are reflected in the mean values shown in Table 1. Of the MCGICP variants, the combined descriptor produced the best results followed by the color descriptor

Table 1
Summary of translation and rotation errors of the methods on the Ford dataset.

	Color ICP	GICP	CCNDT	MCGICP intensity	MCGICP color	MCGICP combined
mean error [m]	0.3532	0.2344	0.2090	0.2370	0.2045	**0.1606**
std. dev. [m]	0.2935	0.2965	0.1952	0.2419	0.1943	**0.1492**
mean error [rad]	0.0030	0.0042	0.0052	0.0035	0.0033	**0.0030**
std. dev. [rad]	0.0021	0.0034	0.0032	0.0023	0.0022	**0.0020**



(a) Translational error.

(b) Rotational error.

Fig. 4. Comparison of accuracy results for MCGICP against existing algorithms performed on the Ford dataset. Translational (top) and rotational (bottom) errors are shown.

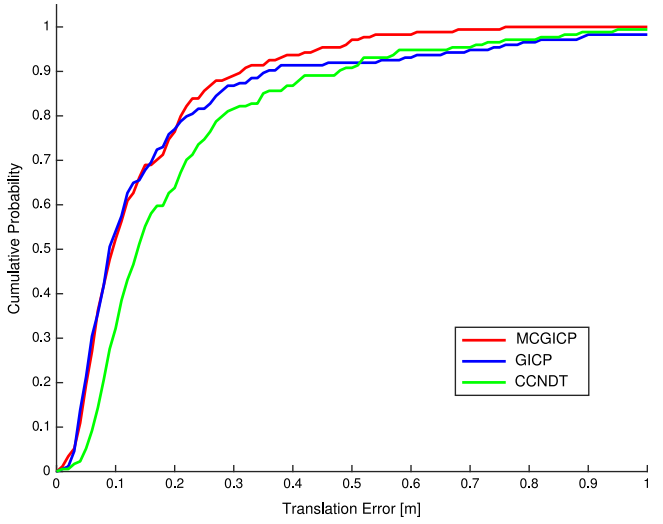


Fig. 5. Cumulative error distribution of the MCGICP method compared to existing methods on the Ford.

while intensity alone produce poor results. This is expected as the combined descriptor provides the most robust information while the intensity alone has minimal distinctive variation in value. The combined descriptor increases the information at each point while maintaining consistency. This results in increased correlation and correspondence certainty. This shows that the MCGICP method is dependent on the use of an accurate and discriminating descriptor space.

The cumulative distribution of error is presented in Fig. 5. The distribution shows that for MCGICP 90% of registrations are performed with less than 0.32 m of error. MCGICP demonstrates improved performance relative to the other methods over the majority of cases, and with slightly improved accuracy over GICP overall. The distribution for MCGICP shows that the method maintains good performance over the entire range of data. Color ICP is omitted from this plot for clarity, due to its failure to converge from the initial position in many cases.

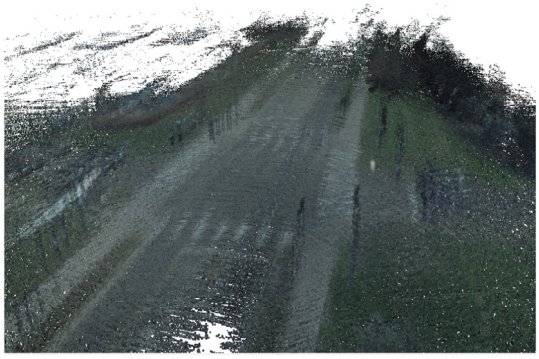
The aggregated maps of a challenging section of the Ford dataset generated by Color ICP, GICP and MCGICP using the combined descriptor are shown in Fig. 6. In the aggregate maps, it can be seen that MCGICP creates more accurate results. This is evident by the blurring which can be seen in the Color ICP and GICP maps but is reduced in the MCGICP results.

An aggregate map of a large section of the Ford dataset generated using scan-to-map matching with the MCGICP algorithm is shown in Fig. 7. Only a subsection of the map, aggregating every 20th scan, is shown for visual clarity. The map shows a highly consistent path, with only minimal drift throughout the loop. The use of scan-to-map matching, where scans are matched to an aggregation of previous scans, improves the global consistency of the results. Results could be further refined using a standard back-end optimization tool such as g2o [30].

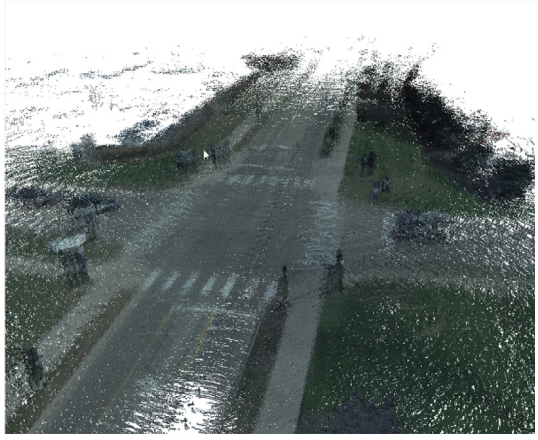
8.2. Freiburg RGBD indoor dataset

The Freiburg RGBD dataset used for these experiments is the *Long Office Household Scene*. In this case, since the dataset only contains points with color information, only color is used as the descriptor channel. This dataset consists of 162 scans with ground truth information from a motion capture system. Scan-to-scan matching is performed and compared to the ground truth information to evaluate accuracy. No pose prior was used, so the initial estimate for the optimization was set to zero in all cases.

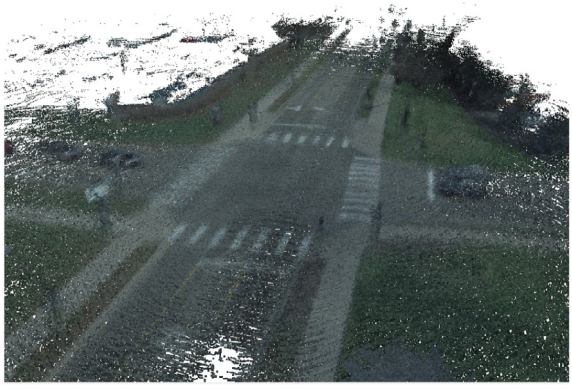
The ideal parameters used for each method were again determined experimentally in order to produce the best results for this dataset. The optimization parameters are the same as the previous



(a) Color ICP.



(b) GICP.



(c) MCGICP combined.

Fig. 6. Aggregated point cloud maps generated from a subsection the Ford Vision and Lidar dataset.

tests. GICP used a max correspondence distance of 0.2 m and MCGICP used a correspondence distance of 10 in the multi-channel descriptor space. NDT used a cell size of 0.05 m and CCNDT uses a color clustering threshold of 5 and a distance threshold of 0.05 with a minimum cluster size of 3. MCGICP used a weighting vector for the nearest neighbor search of $\alpha = \{\alpha_x, \alpha_y, \alpha_z, \alpha_R, \alpha_G, \alpha_B, \alpha_e\} = \{1.0, 1.0, 1.0, 0.02, 0.02, 0.02, 1.0\}$. Finally, MCGICP used an idealized measurement model such that the diagonal of Λ was set to $\{50, 50, 50\}$ corresponding to the RGB channels respectively with zero cross correlation.

Table 2 shows the mean and standard deviation of the errors of MCGCIP compared to the other methods. It shows that on the Freiburg dataset, MCGICP has superior accuracy in both translation and rotation, and with a tighter standard deviation than the other methods.

Table 2
Summary of translation and rotation errors of the methods on the Freiburg dataset.

	NDT	GICP	CCNDT	MCGICP color
mean error [m]	0.0546	0.0528	0.0413	**0.0353**
std. dev. [m]	0.0592	0.0616	0.0511	**0.0415**
mean error [rad]	0.0676	0.0460	0.0408	**0.0349**
std. dev. [rad]	0.0998	0.0625	0.0498	**0.0412**

Fig. 8 shows the accuracy comparison of MCGICP with existing methods. The graph shows that both MCGICP and CCNDT have improved accuracy over the other methods, however in this case MCGICP and CCNDT have approximately equivalent results. This is because this scene is ideally suited for CCNDT as it has high variation in color. MCGICP is also able to use this color information effectively but not to the extent that it improves in accuracy significantly over CCNDT. It should be noted however that MCGICP does show fewer outliers in the registration results than CCNDT, as seen by the error means in Table 2, demonstrating that MCGICP is more robust even in this case.

The cumulative distribution of errors generated using the Freiburg dataset is shown in Fig. 9. The distribution shows that 90% of MCGICP registrations have errors less than 0.067 m. The distribution also shows MCGICP has improved accuracy over both GICP and NDT over the entire range was well as over CCNDT in the top 20th percentile. This demonstrates the robustness of the MCGICP method is it maintains high accuracy over the largest range of possible scans.

8.3. Kinect sparse geometry data

The Kinect Sparse Geometry dataset was obtained using the Microsoft Kinect RGB-D sensor mounted to a mobile robotics test platform. The sequence is of a flat wall which is covered in posters. This sequence contains no distinct geometric surfaces other than the flat wall and therefore produces degenerate solutions in the x and y directions when only geometry is considered. The sequence consists of 200 frames traversing from right to left over a 10 m^2 area. The aggregated results of pairwise registration for MCGICP, GICP and Color ICP are presented in Fig. 10. This experiment employs the same algorithm parameters as were used in the Freiburg dataset.

Fig. 10 demonstrates the ability of the MCGICP algorithm to use the additional channels to compensate for the lack of geometric information. CCNDT performs comparably to MCGICP in this case and is not shown. Color ICP is also able to partially compensate but has much lower accuracy observed by the increased blurring of the posters. Color ICP has trouble correctly aligning the scans due to noise in the images causing incorrect correspondences and skewing the results. The GICP results fail catastrophically in this case due to the lack of geometric information along the wall and therefore all the scans are incorrectly aligned.

8.4. Convergence rate

The overall computation time of MCGICP was compared against the GICP and CICP methods and results are shown in Fig. 11. The box plot shows that MCGICP has a slightly lower median computation time than the other methods as well as a low standard deviation than CCNDT and GICP. Overall MCGICP demonstrates comparable or slightly superior performance over existing scan matching algorithms.

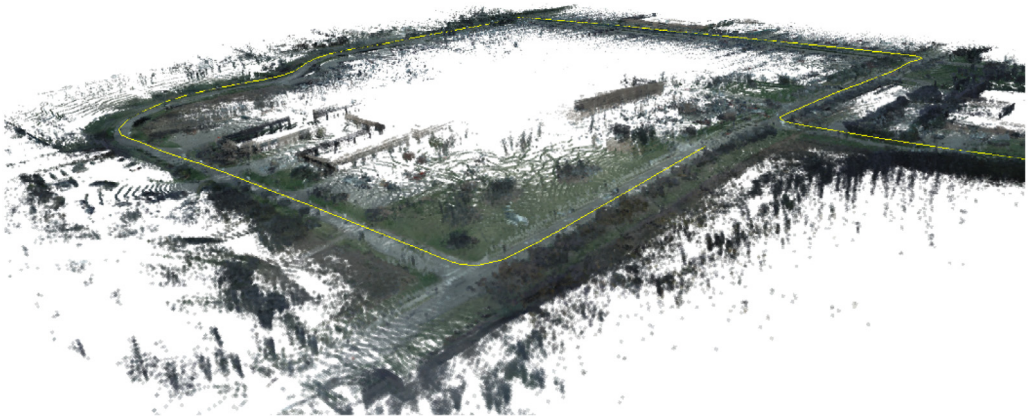
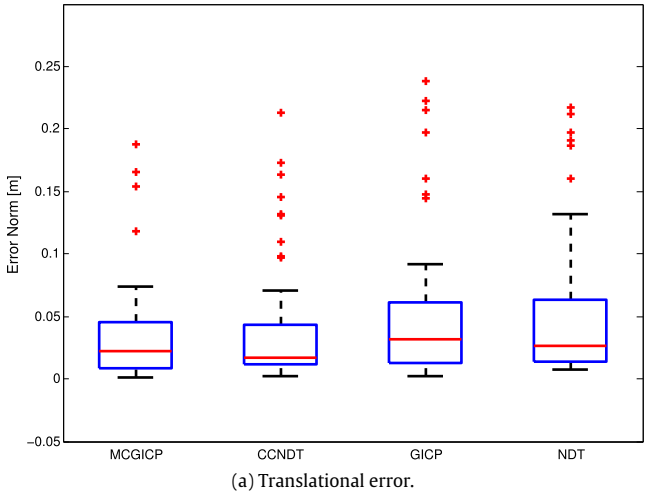
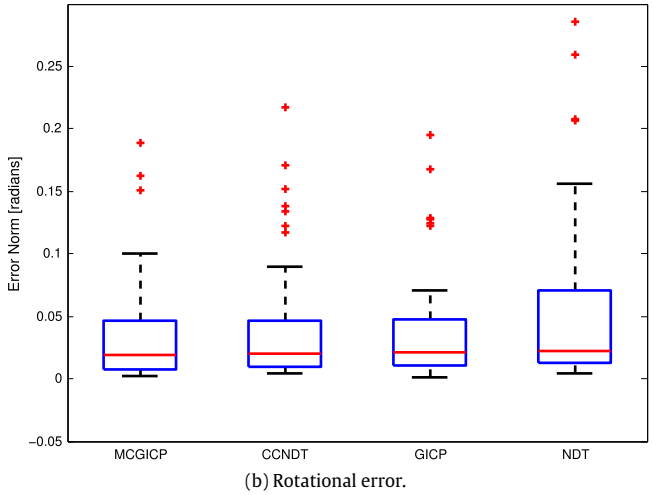


Fig. 7. Aggregate map of a subsection of the Ford dataset generated using MCGICP with scan-to-map matching. The yellow line indicated the calculated path of the vehicle.



(a) Translational error.



(b) Rotational error.

Fig. 8. Comparison of accuracy results for MCGICP against CCNDT and existing algorithms performed on the Freiburg dataset.

The final evaluation compares the convergence rates of the three algorithms. Given an example frame from the Ford Data Set, the error residual is plotted versus the iteration in Fig. 12. As all three algorithms use computationally similar cost functions, iterations are proportional to convergence time.

Fig. 12 clearly shows that MCGICP converges significantly more quickly than the original GICP algorithm. Color ICP has been shown

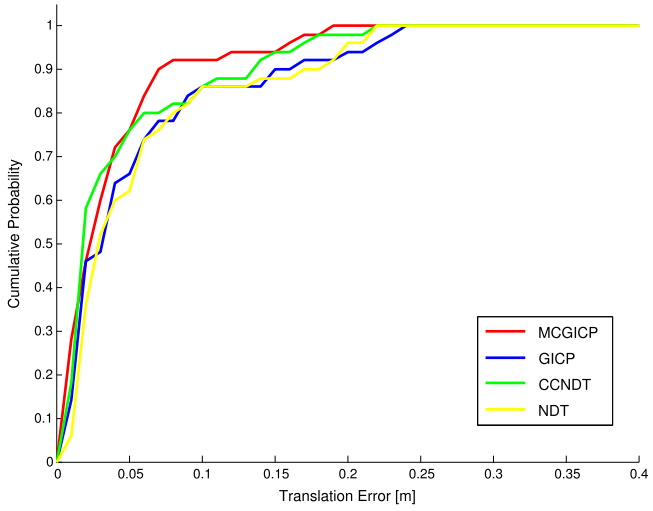


Fig. 9. Cumulative error distribution of the MCGICP method compared to existing methods on the Freiburg dataset.

to converge more quickly than standard ICP due to the fact that it acquires the correct correspondences more quickly using the higher dimensional search space. GICP is shown to converge more quickly than Color ICP, as GICP does not rely as heavily on correct point correspondences to converge and only needs corresponding points to lie on the same surface. MCGICP combines the beneficial properties of both GICP and CICP to acquire the correct surface correspondences and converges most rapidly.

9. Conclusion

This work presents the Multi Channel Generalized-ICP method for robust scan matching. The proposed method incorporates the additional sensor channels directly into the GICP formulation to provide additional information in the plane parallel to the local surface. The GICP algorithm relies solely on surface normal information at each point, and requires surface normals from the point set to span all of $\mathbb{R}^3$ to properly determine the transformation and avoid degeneracy of the solution. However, many real world environments, such as hallways and flat open spaces, do not provide sufficient information when using surface normals alone. The MCGICP method modifies the model covariance planar to the surface normal and calculates nearest neighbor correspondences in a higher dimensional space, thereby exploiting the additional information available in the point cloud from secondary sensor

(a) GICP.

(b) Color ICP.

(c) MCGICP.

Fig. 10. Aggregated point cloud maps generated from the geometrically degenerate poster wall dataset using the Kinect sensor.

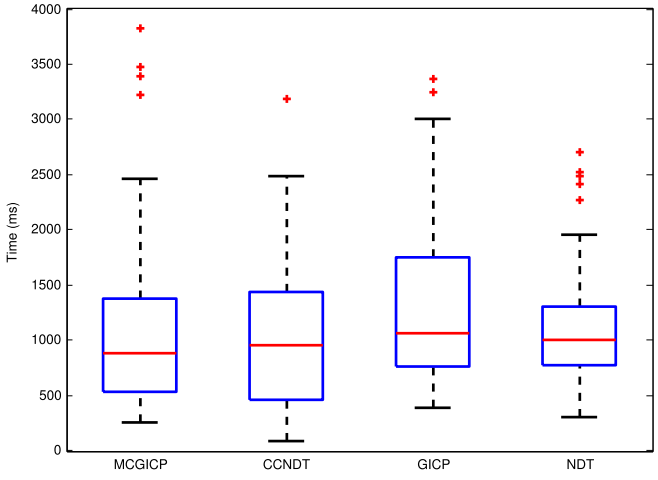


Fig. 11. Comparison of computation time of MCGICP against existing methods.

channels to avoid this shortcoming. The proposed method demonstrates improved registration accuracy and convergence rate on as well as robustness to degenerate geometric cases. MCGICP can perform robustly in both indoor and outdoor environments and with a wider range of possible configurations than previous methods.

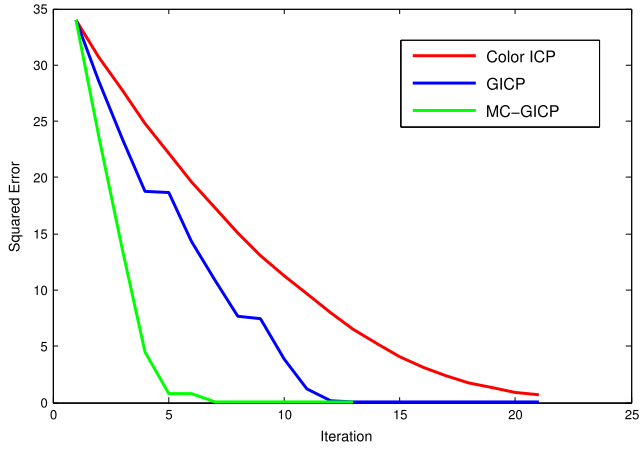


Fig. 12. Comparison of convergence of the Color ICP (red), GICP(blue) and MCGICP (green) algorithms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Future work includes incorporating the proposed method with a graph-SLAM back-end for loop closure and exploring other possible descriptor space combinations. Also, further investigation in the use of intensity and color information together and the effects of lighting and reflectance changes, such as those discussed in [4], needs investigation. Finally, characterization of the domain of registration convergence is an area of future investigation.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) (150692565) through the NSERC Canadian Field Robotics Network (NCFRN), and by a NSERC PGS M scholarship.

References

[1] P. Besl, H. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.
[2] A. Segal, D. Haehnel, S. Thrun, Generalized-ICP, in: Robotics: Science and Systems, RSS, Seattle, USA, 2009, pp. 26–27.
[3] P. Biber, W. Strasser, The normal distributions transform: a new approach to laser scan matching, in: IEEE International Conference on Robotics and Automation, ICRA, vol. 3, IEEE, 2003, pp. 2743–2748.
[4] F. Pomerleau, F. Colas, R. Siegwart, S. Magnenat, Comparing ICP variants on real-world data sets, Auton. Robots 34 (3) (2013) 133–148.
[5] A.E. Johnson, S. Bing Kang, Registration and integration of textured 3D data, Image Vis. Comput. 17 (2) (1999) 135–147.
[6] C Schutz, T. Jost, H Hugli, Multi-feature matching algorithm for free-form 3D surface registration, in: Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on, vol. 2, IEEE, 1998, pp. 982–984.
[7] S. Druon, M.-J. Aldon, A. Crosnier, Color constrained icp for registration of large unstructured 3d color data sets, in: Information Acquisition, 2006 IEEE International Conference on, IEEE, 2006, pp. 249–255.
[8] G. Godin, M. Rioux, R. Baribeau, Three-dimensional registration using range and intensity information, in: Photonics for Industrial Applications, International Society for Optics and Photonics, 1994, pp. 279–290.
[9] H. Yoshitaka, K. Hirohiko, O. Akihisa, Y. Shin'ichi, Mobile robot localization and mapping by scan matching using laser reflection intensity of the sokuiki sensor, in: IEEE Industrial Electronics, IECON 2006-32nd Annual Conference on, IEEE, 2006, pp. 3018–3023.
[10] B. Huhle, M. Magnusson, W. Straßer, A.J. Lilienthal, Registration of colored 3D point clouds with a kernel-based extension to the normal distributions transform, in: IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2008, pp. 4025–4030.
[11] J. Servos, S.L. Waslander, Multi channel generalized-ICP, in: Robotics and Automation, ICRA, 2014 IEEE International Conference on, IEEE, 2014, pp. 3644–3649.
[12] G. Pandey, J. McBride, R. Eustice, Ford campus vision and lidar data set, Int. J. Robot. Res. 30 (13) (2011) 1543–1552.

[13] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers, A Benchmark for the Evaluation of RGB-D SLAM Systems, in: Proc. of the International Conference on Intelligent Robot Systems, IROS, 2012.
[14] Y. Chen, G. Medioni, Object modeling by registration of multiple range images, in: IEEE International Conference on Robotics and Automation, ICRA, vol. 3, 1991, pp. 2724–2729.
[15] M. Magnusson, T. Duckett, A. Lilienthal, Scan registration for autonomous mining vehicles using 3D-NDT, J. Field Robot. 24 (10) (2007) 803–827.
[16] T. Stoyanov, M. Magnusson, A. Lilienthal, Point set registration through minimization of the L2 distance between 3D-NDT models, in: IEEE International Conference on Robotics and Automation, ICRA, St. Paul, MN, USA, 2012, pp. 5196–5201.
[17] J. Servos, S.L. Waslander, Using RGB information to improve NDT distribution generation and registration convergence, in: International Conference on Intelligent Unmanned Systems, ICIUS, Montreal, Canada, 2014.
[18] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, W. Burgard, An evaluation of the RGB-D SLAM system, in: IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2012, pp. 1691–1696.
[19] K. Konolige, M. Agrawal, J. Sol, Large-Scale Visual Odometry for Rough Terrain, in: M. Kaneko, Y. Nakamura (Eds.), Robotics Research, in: Springer Tracts in Advanced Robotics, vol. 66, Springer Berlin, Heidelberg, 2011, pp. 201–212.
[20] G. Pandey, J. McBride, S. Savarese, R.M. Eustice, Visually bootstrapped generalized ICP, in: IEEE International Conference on Robotics and Automation, ICRA, 2011, pp. 2660–2667.
[21] B. Huhle, P. Jenke, W. Straßer, On-the-fly scene acquisition with a handy multisensor system, Int. J. Intell. Syst. Technol. Appl. 5 (3) (2008) 255–263.
[22] Nvidia, CUDA: Compute unified device architecture programming guide, Santa Clara, CA, 2007.
[23] R.A. Newcombe, S. Lovegrove, A. Davison, DTAM: Dense tracking and mapping in real-time, in: IEEE International Conference on Computer Vision, ICCV, 2011, pp. 2320–2327.
[24] R. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohli, J. Shotton, S. Hodges, A. Fitzgibbon, KinectFusion: Real-time dense surface mapping and tracking, in: IEEE International Symposium on Mixed and Augmented Reality, ISMAR, Basel, Switzerland, 2011, pp. 127–136.
[25] T. Whelan, J. McDonald, M. Kaess, M. Fallon, H. Johannsson, J.J. Leonard, Kintinuous: Spatially Extended KinectFusion, in: RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras, 2012.
[26] M. Muja, D.G. Lowe, Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration, in: International Conference on Computer Vision Theory and Application, VISSAPP, INSTICC Press, 2009, pp. 331–340.
[27] P. Corke, R. Paul, W. Churchill, P. Newman, Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2013, pp. 2085–2092.
[28] G. Pandey, J.R. McBride, S. Savarese, R. Eustice, Automatic targetless extrinsic calibration of a 3D lidar and camera by maximizing mutual information, in: Association for Advancement of Artificial Intelligence, AAAI, Toronto, Canada, 2012.
[29] R.B. Rusu, S. Cousins, 3D is here: Point Cloud Library (PCL), in: IEEE International Conference on Robotics and Automation, ICRA, Shanghai, China, 2011.
[30] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, W. Burgard, g2o: A general framework for graph optimization, in: IEEE International Conference on Robotics and Automation, ICRA, Shanghai, China, 2011.

James Servos received his B.S. in 2012 and his M.S. in 2014 from the University of Waterloo in the department of Mechanical and Mechatronics Engineering. He is currently employed by Clearpath Robotics as an Autonomy Engineer, and is actively involved in the deployment of SLAM and scan registration methods for both research and industrial applications.

Steven L. Waslander received his B.Sc.E. in 1998 from Queen's University, his M.S. in 2002 and his Ph.D. in 2007, both from Stanford University in Aeronautics and Astronautics. He was a Control Systems Analyst for Pratt & Whitney Canada from 1998 to 2001. In 2008, he joined the Department of Mechanical and Mechatronics Engineering at the University of Waterloo in Waterloo, ON, Canada where he is now an associate professor. He is the Director of the Waterloo Autonomous Vehicles Laboratory (WAVELab, http://wavelab.uwaterloo.ca). His research interests are in the areas of autonomous aerial and ground vehicles, simultaneous localization and mapping, nonlinear estimation and control, and multi-vehicle systems.