Registration of Colored 3D Point Clouds with a Kernel-based Extension to the Normal Distributions Transform

Benjamin Huhle, Martin Magnusson, Wolfgang Straßer and Achim J. Lilienthal

***Abstract*— We present a new algorithm for scan registration of colored 3D point data which is an extension to the Normal Distributions Transform (NDT). The probabilistic approach of NDT is extended to a color-aware registration algorithm by modeling the point distributions as Gaussian mixture-models in color space. We discuss different point cloud registration techniques, as well as alternative variants of the proposed algorithm. Results showing improved robustness of the proposed method using real-world data acquired with a mobile robot and a time-of-flight camera are presented.**

I. Introduction

Registration of 3D point clouds is an essential part of many applications. In robotics, many of the common Simultaneous Localization And Mapping (SLAM) solutions rely on the registration of point clouds. For object detection and visualization, e.g., the 3D point cloud is typically textured. However, during the model building color data is often used in the texturing step only.

Enhancing the information by adding color in an early stage of the pipeline should help to increase registration quality and robustness. Approaches where visual cues are employed for registration are often based on local features in images that are mapped to single points in the range data. Using these small samples of the geometry for matching can lead to significant errors, especially when using noisy or erroneous depth sensors. Therefore, color data should be incorporated also in registration algorithms that make use of all available geometric information and which can then be combined with feature-based methods if needed.

We present an extension to the Normal Distributions Transform (NDT, see section II-C) which adds color information to NDT's probabilistic local surface models.

II. Overview of registration algorithms

The goal of scan registration is to find a rigid-body transformation T that minimizes a certain error function between a *data set* X and another set of points Y (the *model set*), where both point sets can be partially or fully overlapping. The following three algorithms can be classified as local registration methods, which means that they search for the closest local optimum. Therefore, they rely on a good estimate of the initial pose in order to converge to the correct solution.

B. Huhle and W. Straßer are with the Department of Graphical Interactive Systems WSI/GRIS, University of Tübingen, Germany.
huhle@gris.uni-tuebingen.de,
strasser@gris.uni-tuebingen.de

M. Magnusson and A. Lilienthal are with the Center of Applied Autonomous Sensor Systems, Dept. of Technology, Örebro University, Sweden.
martin.magnusson@tech.oru.se,
achim@lilienthals.de

A. Iterated Closest Point (ICP)

The ICP algorithm is widely used today for registration of 3D point clouds. ICP iteratively updates and refines the relative pose by minimizing the sum of squared distances between corresponding points in the two scans. The two seminal papers on ICP are by Besl and McKay [3] and Chen and Medioni [5]. Since its conception, a large number of variants have been developed, and a good survey of different variations of ICP was presented by Rusinkiewicz [16].

At each iteration of the ICP algorithm, a set of point pairs is built, where each point in the data scan is associated with its closest neighbor in the model. A closed-form solution exists for finding the transformation that minimizes the sum of distances between associated pairs. Although the closest point does not generally correspond to the same point on the scanned surface, especially if the scans are far apart from each other, successive iterations will still usually converge to a good solution. However, for only partially overlapping scans, special care has to be taken in order to discard correspondences that connect points in the border regions that do not have a matching point in the model at all. Since this is not handled inherently by the ICP algorithm, such correspondences would otherwise introduce a shift towards a completely overlapping solution.

The two main problems of ICP are that it is point-based, and as such does not make use of the local surface shape around each point, and that the nearest-neighbor search in the central loop is rather time consuming.

B. Color-ICP

The ICP algorithm can naturally be extended to colored data by measuring the distance between corresponding points between the six-dimensional color-space vectors [11]. When using that algorithm, one needs to pay attention to the scaling of the feature elements, depending on the sampling distribution of scan points. For example, if we assume the RGB components of the features have range $[0, 1]$ and the spatial features are measured in centimeters, the colors will have little influence on the result in large-scale environments. With different scaling, or different sample distributions, points with similar colors will be preferred over spatially proximate points. This problem is especially pronounced for data where the scan points are unevenly distributed.

Fig. 1. Feature-based registration in a dynamic environment. Left: Corresponding SIFT-features. Right: Resulting model.

C. The Normal Distributions Transform (NDT)

The normal distributions transform method for registration of 2D data was introduced by Biber and Straßer [4]. The key element in this algorithm is a new representation for the model point cloud. Instead of matching the data point cloud to the points in the model directly, the probability of finding a point at a certain position is modeled by a linear combination of normal distributions. This gives a piecewise smooth representation of the model surface, with continuous first and second order derivatives. Using this representation, it is possible to apply standard numerical optimization methods for registration. Because the points in the target scan are not used directly for matching, there is no need for computationally expensive nearest-neighbor search, as in ICP. Computing the normal distributions is a one-off task that is done during a single pass through the points of the target scan.

NDT has been extended to 3D and shown to perform well in comparison with ICP [14]. In particular, the normal distributions transform is insensitive to uneven sample distributions, and does not employ an explicit nearest neighbor search. When using NDT for localization and mapping, the original point clouds do not need to be stored, and the NDT representation efficiently compresses the scan data, largely reducing the memory requirements for large-scale maps.

D. Local feature based registration

We now discuss a number of approaches that, in contrast to the methods of the preceding section, use more salient features of the scene for registration.

Good local surface descriptions should be invariant to rigid motion (that is, rotation and translation), so that corresponding surface parts can be found regardless of the initial poses of the scans. If sufficiently discriminative features can be found, the need for a good initial pose estimate is alleviated.

Several ways of describing local surface shape have been proposed, such as spin-images [12] and surface signatures [17]. A common weakness of these methods is that they rely on *oriented points* — that is, points with a surface normal. If the surface normal cannot be accurately computed, as is the case for noisy data, the local surface descriptions are not useful for finding corresponding surface patches.

When images are available in addition to range data, point clouds can be registered by relying on the local visual structure of the scene for solving the data association (correspondence) problem using visual features (e.g., [2]).

Since feature-based methods rely on a small number of 3D points, instead of using all the available data, noisy range readings can cause significant errors. Similarly, a single false correspondence can lead to severe misalignment even in the presence of several correct correspondences. Even though the state of the art Scale-Invariant Feature Transform (SIFT, [13]) features are generally very robust, false correspondences can occur if part of the model has a repetitive texture, for example. In a dynamic environment, a small modification can cause these methods to fail (cf. Figure 1) as well, even if the scene remains the same on a larger scale. The methods discussed in Sections II-A–II-C, as well as the proposed Color-NDT, however, would yield correct results.

E. Combined Features/NDT registration

In order to reduce the effect that noisy range data have on feature-based registration, a method was proposed by Huhle et al. [10] that uses an energy function

$$E = \alpha E_{NDT} + (1 - \alpha) E_F, \tag{1}$$

which combines the NDT score function E_{NDT} and an energy function E_F that penalizes distances of corresponding color features. The weight α is determined by the result of the preceding rough alignment using the same feature correspondences. Scans captured by sensors with a narrow field of view often have limited geometric structure, in which case NDT can fail. Therefore, α should favor the feature solution for such scenes.

III. Color-NDT

NDT has a number of valuable properties and has been shown to perform well in comparison to standard registration methods. In order to increase the robustness, especially for scenes with few geometric constraints, we extend NDT into a color-aware registration method to gain from the additional sensor information.

Note that another approach to use NDT in the color domain has been presented by [1]. That method is also called Color-NDT. However, it is meant for change detection and cannot be used for registration applications.

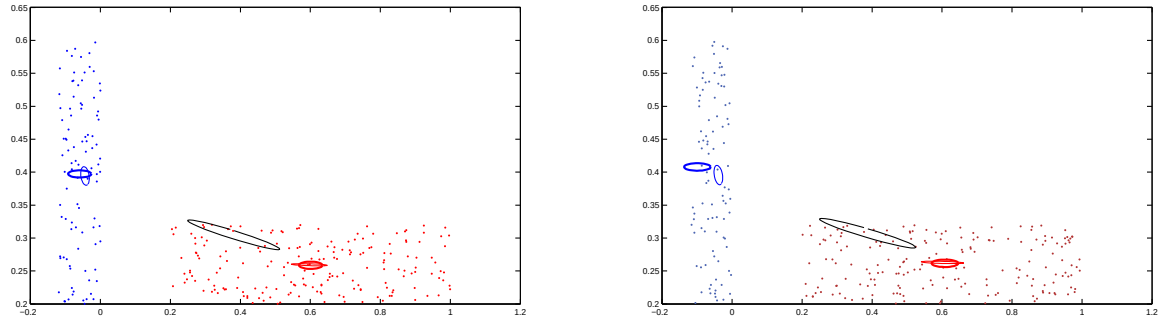

Fig. 2. This figure shows one cell with conditional distributions of 6D-NDT (thick blue/red), distribution of 3D-NDT (black), and of Color-NDT with adaptive kernels (thin blue/red). For visualization, dimensionality has been reduced to 2D space and color consists of only the hue channel. Left: conditional distribution/kernel for exactly matching colors. Right: The colors of the data and the model are slightly different.

A. Color-NDT using adaptive kernels

In order to incorporate color information into the NDT registration, we represent the point distribution in each cell of the NDT grid as a mixture model in color space. (In the following we use the notation x_i for spatial coordinates and $\tilde{x}_i$ for color coordinates of data points.) For each cell k we build the mixture model of the color distribution

$$p_k(\tilde{x}) = \sum_{j=1}^{M} \alpha_j \mathcal{N}(\tilde{q}_{k,j}, \tilde{\mathbf{\Sigma}}_{k,j}),$$

employing M components with means $\tilde{q}_{k,j}$ and covariances $\tilde{\mathbf{\Sigma}}_{k,j}$.

This mixture density is estimated with the Expectation Maximization algorithm (EM, [6]) using the color space coordinates $\tilde{y}_i$ of the points belonging to the reference scan. We apply the EM algorithm for maximum likelihood estimation of mixture densities as described in [15]. The initial guesses of the component distributions are determined via k-means [8].

The next step is to build a Gaussian Mixture Model (GMM) of the point distributions in each cell. The components of the color space model are used as kernel functions centered on their means in color space. The kernels weight the influence of the points when building the spatial model. Thus, for each kernel j, we get a corresponding component j of the GMM. Accordingly, the color weights

$$\xi_{ij} = \exp\left(-\frac{1}{2}(\tilde{y}_i - \tilde{q}_{c(i),j})^t \tilde{\mathbf{\Sigma}}^{-1}_{c(i),j}(\tilde{y}_i - \tilde{q}_{c(i),j})\right) \quad (2)$$

are determined by evaluating the color kernels. The index $c(i)$ here points to the cell where data point y_i falls into.

Building the model for the Color-NDT cells is done by computing the weighted spatial means

$$q_{k,j} = \frac{1}{\Xi_{k,j}} \sum_{i=1}^{N} \delta(c(i), k)\ \xi_{ij}\ y_i \quad (3)$$

and weighted spatial covariances

$$\mathbf{\Sigma}_{k,j} = \frac{\Xi_{k,j}}{\Xi_{k,j} - \sum_i \delta(c(i), k)\ \xi^2_{k,j}} \quad (4)$$

$$\sum_{i=1}^{N} \xi_{ij}(y_i - q_{c(i),j})(y_i - q_{c(i),j})^t \quad (5)$$

for each cell k, where δ denotes the Kronecker delta. In Equations 3–5, the sum of all color weights for cell k and mixture component j,

$$\Xi_{k,j} = \sum_{i=1}^{N} \delta(c(i), k)\ \xi_{ij},$$

is used for normalization. A visualization of the spatial distributions computed by Color-NDT can be seen in Figure 5. Note that differences in color (brightness) appear due to the artificial illumination of the rendering of the ellipsoids.

The registration algorithm employing the necessary adaptations for solving the presented color-aware version of NDT is as follows: In order to register a new scan to the Color-NDT of the model point cloud, the score s depending on the transformation $T_{\mathbf{p}}$ is to be optimized. This transformation $T_{\mathbf{p}}$ in our case is a rigid-body transformation parametrized by a 6D parameter vector $\mathbf{p}$. Given a sample of N points belonging to the new scan, we compute the score

$$s(\mathbf{p}) = \sum_{i=1}^{N}\sum_{j=1}^{M} \xi_{ij} \exp(-\frac{1}{2}(x'_i - q_{c(i),j})^t \mathbf{\Sigma}^{-1}_{c(i),j}(x'_i - q_{c(i)j})),$$

where the spatial coordinates of the points are denoted by $x'_i = T_{\mathbf{p}}(x_i)$.

The optimization of the score function with regard to the transformation parameters can be done with arbitrary numerical optimization methods. We experienced fast convergence with Newton's method incorporating a line search.

We optimize iteratively over transformation $T_{\mathbf{p}}$, which consists of a translation and a rotation that is parameterized as a rotation vector [7]. Assuming only small angular differences for each step of the iteration, we use the small angle approximation for the first partial derivatives of T. For the second partial derivatives, this leads to an all-zero Hessian matrix which reduces the computational cost computing the second partial derivative of the score function $s(\mathbf{p})$.

Note that, usually, a critical point in mixture density estimation is the choice of the number of components M that are used to represent the density. However, our main concern is not to build a highly accurate model in color space. As in standard 3D-NDT, where the point distributions mainly

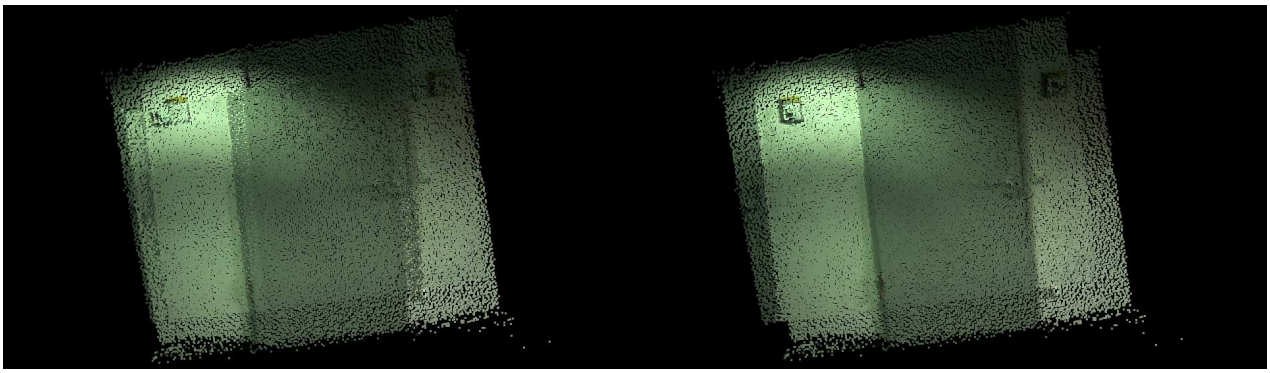

Fig. 3. Door data set (2 frames). Left: initial pose of both frames. Right: registered with Color-NDT.

distinguish differently oriented planar structures, we are interested in a color-space model that distinguishes different colors and enables us, intuitively speaking, to draw the points to be registered into the correct direction depending on their color. Since we are dealing with 3D spatial data it then suffices to compute a 3-components mixture model for each cell in order to fix the registration result to a unique pose even if the actual density in color space is more complex.

Based on these considerations we tried different approximations for representing the color space density which demand less computational effort. One could use kernels with fixed mean and fixed variance. However, as can be expected, this method is less accurate since the resulting distributions are not as expressive as our proposed method. Another approach is to estimate discrete kernels by applying k-means clustering in color space only. However, this approach suffers from discretization effects. Computing an (isotropic) variance in color space from the clustering result and applying this solution as weighting kernels in the spatial domain shows decreased performance compared to the EM estimated kernels version as well.

B. NDT using combined color-space distributions (6D-NDT)

An alternative method for fusing color and range data for NDT is to discretize only along the spatial dimensions, as for standard 3D-NDT, and store six-dimensional normal distributions over the combined 6D color-space feature vectors in each cell. This is the most straight-forward analogon to the Color-ICP version introduced by Johnson and Kang [11]. Building such 6D structures is faster than finding the color kernels as described in section III-A. Optimizing the score is also faster because only one distribution needs to be evaluated for each point in the data scan.

However, in many cases, a single normal distribution is not a good model for the color-space distribution of points. To get a better understanding of this six-dimensional representation, and in order to compare the 6D-NDT with the above described kernel-based Color-NDT, we investigate the conditional distributions of the 6D-NDT, i.e., the spatial distribution given a certain color. We compute the conditional means

$$\mu_k(x|\tilde{x}) = q_k + \mathbf{\Sigma}_k^{x,\tilde{x}} \tilde{\mathbf{\Sigma}}_k^{-1} (\tilde{x} - \tilde{q}_k)$$

and conditional covariances

$$\mathbf{\Sigma}_k(x|\tilde{x}) = \mathbf{\Sigma}_k - \mathbf{\Sigma}_k^{x,\tilde{x}} \tilde{\mathbf{\Sigma}}_k^{-1} \mathbf{\Sigma}_k^{\tilde{x},x},$$

where $(q_k, \tilde{q}_k)$ is the 6D mean and $\mathbf{\Sigma}_k$ and $\tilde{\mathbf{\Sigma}}_k$ denote the covariances in the spatial and the color subspaces of cell k, respectively. Analogously, $\mathbf{\Sigma}_k^{x,\tilde{x}}$ and $\mathbf{\Sigma}_k^{\tilde{x},x}$ denote the cross-covariances in the color and spatial subspaces.

A visualization of the resulting distributions is given in Figure 2. Whereas points that meet the color of the model points exactly (left) are attracted to the correct spatial position, a blue point in the right subplot of Figure 2 is expected by 6D-NDT to lie even further on the left compared to the almost-blue model points. Contrarily, the kernel-based Color-NDT handles this case well, expecting a point with color different from the model's colors to lie closer to the overall (standard NDT) mean.

IV. Experiments

A. Sensor setup

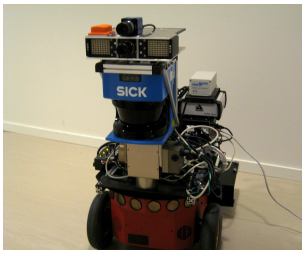

Fig. 4. The sensor setup used in the experiments. The PMD camera is mounted on top of a SICK laser scanner that was not used in this work.

Data were collected using the experimental robot platform *Tjorven*, a Pioneer P3-AT equipped with an Amtec pan/tilt unit (see Figure 4). Range and color images were acquired with a combination of a *PMD vision 19k* time-of-flight camera and a *Matrix-Vision Blue Fox* color camera, mounted on the pan/tilt unit. The data from the two cameras are combined as described in [9]. The time-of-flight camera illuminates the scene with modulated near-infrared light using an array of LEDs. It both measures the reflectance and computes range values for each pixel of the image based on the phase shift

of the incoming modulated light. The resolution of the gray-scale image and depth map is 160×120 pixels, and the maximum refresh rate for the camera is approximately 15 Hz. The PMD camera is sensitive to the lighting conditions of the scene. The range data are quite noisy and contain frequent outliers, especially at depth discontinuities. Outliers in the resulting colored point clouds due to sensor error are pruned and depth data are smoothed with the methods described in [10]. Still, the noise level is significant.

B. Results

Registration methods that use geometric information only cannot correctly align surfaces that lack prominent spatial structure, such as the two scans shown in Figure 3. The extension to NDT proposed in Section III-A also incorporates knowledge about the surface texture into the registration process, and therefore is able to match both point clouds.

The data set depicted in Figure 5 was recorded while driving the robot platform past the scene, looking sideways. We relied on the robot's odometry for the initial pose estimates of the scans. Even though the data set as a whole contains some geometric features, registration using standard 3D-NDT misaligned several of the partial scans, mainly because of two reasons. Firstly, single frames of the set can suffer from the aperture problem, i.e., not capturing enough structure in one view. Secondly, the surface model also describes the high noise level of the depth sensor. The same problems also affect Color-NDT. However, because of the more descriptive surface representation, it performs significantly better on the same test set. The difference is most obvious around the microwave oven in the upper right corner of the images, where strong contrasts occur on mostly planar surfaces.

We also applied the combined energy function approach (proposed in [10], cf. Section II-E) to a data set collected at the same scene. Due to too large steps between the frames, a global alignment using image features is necessary, i.e., a registration using solely 3D-NDT or Color-NDT is not possible. The additional gain of replacing 3D-NDT with Color-NDT in the combined method is shown in Figure 7. The initial alignments, based only on SIFT features, expose some of the problems of relying on a small set of corresponding points with noisy range data. Additionally, we encountered false feature correspondences due to the repetitive patterns in the scene (cf. Figure 6). We increased the weight α from (1) to better show the influence of NDT. Similarly to the previous experiment, we noticed an improved registration along the normals of the planar structures, though coupled with large offsets along the other directions for standard 3D-NDT. Once again, the improved robustness of Color-NDT shows in a significantly enhanced registration result.

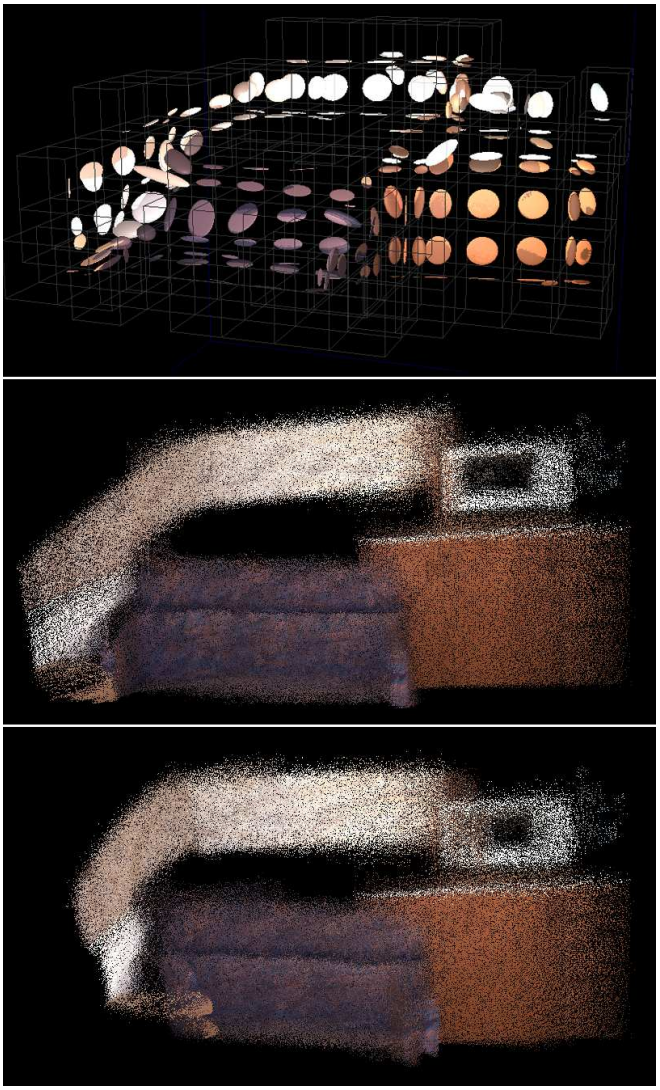

Fig. 5. Data set A (21 frames, sequentially registered in an incremental fashion). Top: illustration of the Color-NDT cell grid. Middle: registration using Color-NDT. Bottom: registration using standard NDT.

Fig. 6. False correspondences in two frames of data set B.

V. SUMMARY

The two main contributions of this paper are, firstly, a discussion of several approaches to scan registration using spatial and partially also color information. Secondly, a new algorithm for scan registration of colored 3D data has been presented. We have demonstrated its increased robustness compared to the original 3D-NDT algorithm as well as to a combined features/NDT method proposed recently. The comparisons were made using real-world data collected with a mobile robot equipped with a PMD time-of-flight camera.

In future work, the matter of adaptive cell splitting should be investigated further.

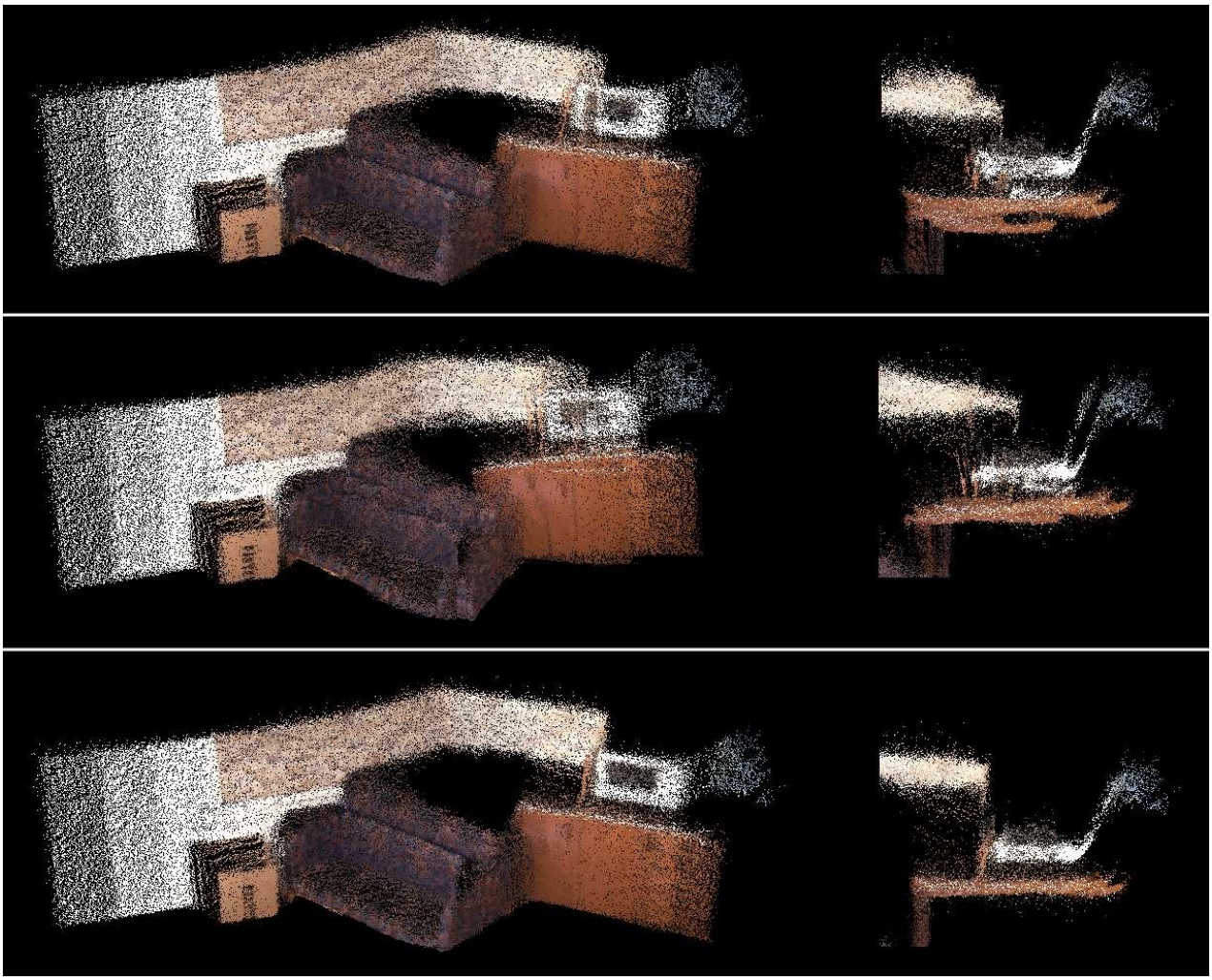

Fig. 7. Data set B (11 frames, sequentially registered in an incremental fashion) with detail views seen from above in the right column. Top: Feature-based registration only. Middle: registration using combined features/NDT method. Bottom: registration using combined features/Color-NDT method. The figure shows the additional gain of replacing 3D-NDT with Color-NDT.

REFERENCES

[1] Henrik Andreasson and Achim Lilienthal. Has Something Changed Here? Autonomous Difference Detection for Security Patrol Robots. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3429–3435, 2007.

[2] Henrik Andreasson and Achim Lilienthal. Vision Aided 3D Laser Scanner Based Registration. In *Proc. 3rd European Conference on Mobile Robots (ECMR)*, pages 192–197, 2007.

[3] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14(2):239–256, 1992.

[4] Peter Biber and Wolfgang Straßer. The Normal Distributions Transform: A New Approach to Laser Scan Matching. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2743–2748, 2003.

[5] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, April 1992.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.

[7] James Diebel. Representing Attitude: Euler Angles, Quaternions, and Rotation Vectors. Technical report, Stanford University, Palo Alto, CA, 2006.

[8] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.

[9] Benjamin Huhle, Sven Fleck, and Andreas Schilling. Integrating 3D Time-of-Flight Camera Data and High Resolution Images for 3DTV Applications. In *Proc. IEEE 3DTV Conference (3DTV CON)*, 2007.

[10] Benjamin Huhle, Philipp Jenke, and Wolfgang Straßer. On-the-Fly Scene Acquisition with a Handy Multi-Sensor System. *Int. J. of Intelligent Systems Technologies and Applications*, to appear, 2008.

[11] A. Johnson and S. Kang. Registration and Integration of Textured 3D Data. In *Proc. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM)*, pages 234–241, 1997.

[12] Andrew Edie Johnson. *Spin Images: A Representation for 3-D Surface Matching*. PhD thesis, Carnegie Mellon University, 1997.

[13] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[14] Martin Magnusson, Achim Lilienthal, and Tom Duckett. Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics*, 24(10):803–827, 2007.

[15] R. Redner and H. Walker. Mixture Densities, Maximum Likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, 1984.

[16] Szymon Marek Rusinkiewicz. Efficient Variants of the ICP Algorithm. In *Proc. International Conference on 3-D Digital Imaging and Modeling (3DIM)*, pages 145–152, 2001.

[17] Sameh M. Yamany and Aly A. Farag. Free-Form Surface Registration using Surface Signatures. In *Prof. IEEE International Conference on Computer Vision (ICCV'99)*, pages 1098–1104, 1999.