Automatic Registration of Multiple Texel Images (fused Lidar/Digital Imagery) for 3D Image Creation

Scott E. Budge, Neeraj Badamikar

Center for Advanced Imaging Lidar, Utah State University,
Logan, UT 84322-4120, (435) 797-3433

ABSTRACT

Creation of 3D images through remote sensing is a topic of interest in many applications such as terrain/building modeling and automatic target recognition (ATR). Several photogrammetry-based methods have been proposed that derive 3D information from digital images from different perspectives, and lidar-based methods have been proposed that merge lidar point clouds and texture the merged point clouds with digital imagery. Image registration alone has difficulty with smooth regions with low contrast, whereas point cloud merging alone has difficulty with outliers and lack of proper convergence in the merging process.

This paper presents a method to create 3D images that uses the unique properties of texel images (pixel-fused lidar and digital imagery) to improve the quality and robustness of fused 3D images. The proposed method uses both image processing and point-cloud merging to combine texel images in an iterative technique. Since the digital image pixels and the lidar 3D points are fused at the sensor level, more accurate 3D images are generated because registration of image data automatically improves the merging of the point clouds, and vice versa. Examples illustrate the value of this method over other methods.

Keywords: lidar, ladar, texel image, texel camera, sensor fusion, point-cloud matching

1. INTRODUCTION

Generation of 3D imagery is a topic of interest in many applications. These datasets result from texturing a 3D surface with digital imagery, resulting in a 3D image made up of textured elements, or "texels," as described in computer graphics literature. The 3D images can then be used to create models of historical sites, document archaeological artifacts, or document topography. In surveillance and defense applications, 3D imagery can be used to identify persons or perform automatic target recognition (ATR).

There have been many efforts to create datasets that include both 3D sensed data and imagery from a scene. These include hardware configuration methods such as combining stereo camera pairs with a lidar sensor,[1–3] methods based on edges, geometric features, or match points,[4–8] methods based in optical flow,[9] or methods based on mutual information.[10]

There have also been many efforts to match 2D datasets and create 3D models from the images. Many of the 2D methods are based in the well known mathematics of projective geometry and stereo pairs.[11–13]

Other efforts have been made in matching 3D datasets. The seminal work of Besl and McKay[14] in developing an iterative method for matching 3D point clouds has led to various improvements to the performance of the basic algorithm.[15] Other 3D methods have been developed that are guided by lidar intensity data.[16]

The goal of this work is to present a method that uses previously fused imagery and lidar data to register 3D textured surfaces. If the datasets are taken from different perspectives around a 3D scene, the registered datasets will produce a true 3D image of the scene. Since the datasets contain fused 2D and 3D data, both 2D and 3D registration methods can be employed to better register the datasets.

The remainder of the paper proceeds as follows. Section 2 begins by describing the combined sensor used in this work. Section 3 describes the basic method used to register the datasets based on the digital imagery, followed by an iterative improvement exploiting the 3D lidar data described in Section 4. Finally, registration results are given in Section 5, and Section 6 concludes the paper.

Figure 1: Texel camera constructed from a TOF depth sensor and an imaging (EO) sensor.
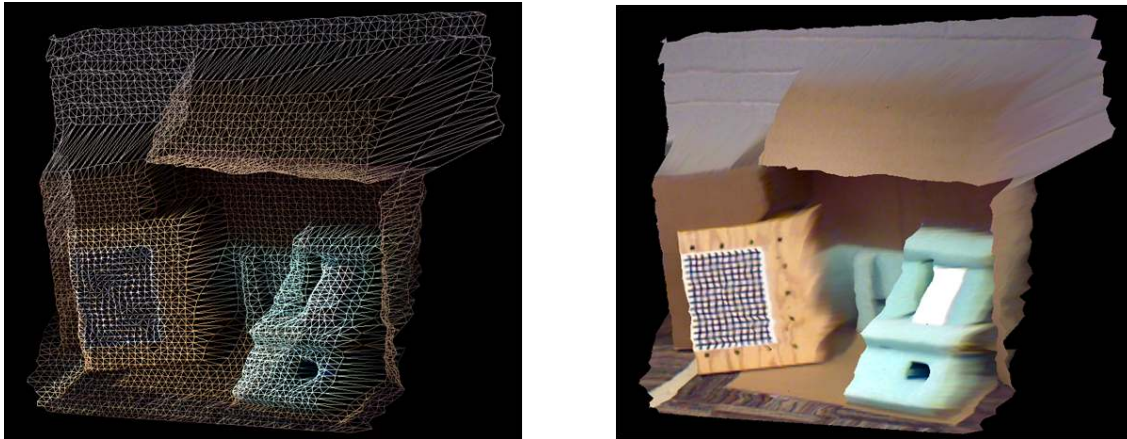
| (a) | (b) |

Figure 2: Examples of a texel image. (a) Triangulated, interconnected network (TIN) created from the lidar measurements. (b) Textured TIN surface, or texel image.

2. THE TEXEL CAMERA

The combination of a flash or TOF lidar with a digital (electro-optic or EO) camera that produces data that is fused at the pixel level is known as a "texel camera."[18–20] An example of a compact texel camera prototype is shown in Figure 1. The camera is constructed to produce fused datasets directly – no post processing is necessary. This is possible because each 3D lidar measurement obtained from the lidar sensor array is assigned, through a calibration process, to a pixel in the simultaneously captured EO image. The EO pixel assigned to each lidar measurement is the projection of the 3D lidar point onto the EO image.

Since the resolution of the EO sensor is much higher than the resolution of the lidar sensor, only a few of the EO pixels correspond directly to projected 3D points measured by the lidar. As a result, other pixels in the EO image must be assigned 3D points by interpolation of the lidar pixels fused to the surrounding EO pixels. When correctly calibrated, there is a one-to-one correspondence between 3D points in the sensor field of view and the pixels in the EO image. This means that a 3D point, referenced to the COP of the lidar sensor, is assigned to each pixel in the EO image.

An example of a dataset produced by a texel camera is given in Figure 2. A triangulated, interconnected network (TIN) is created from the lidar points, and the EO image is used to texture the triangles. The image resulting from the textured elements lead to the name "texel image."

3. IMAGE REGISTRATION USING TEXEL IMAGES

The creation of 3D images results from the registration of several 2.5D texel images acquired from different viewpoints around the 3D object of interest. If each position and pose of a texel camera is known to a high

---
Algorithm 1 Registration of texel images.
---

1. Detect Harris features.

2. Determine putative correspondences.

3. RANSAC.

4. Estimate the optimal fundamental matrix and perform final testing.

5. Point cloud matching using fused EO image/lidar data.

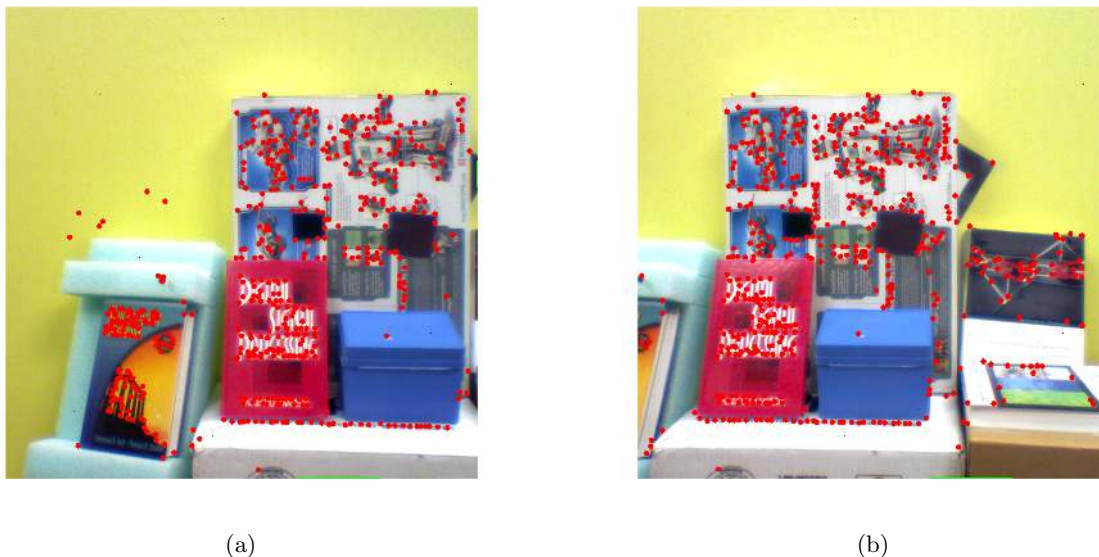(a)   (b)

Figure 3: Examples of Harris features found in the (a) first and (b) second images and marked with red dots.

degree of accuracy in a common frame of reference, it is possible to register the images immediately, and the overlapping areas of the texel images can be combined to create a continuous textured surface. It is of interest, however, to create fused images without the cost of a highly accurate position/pose measurement system.

The fused nature of texel images allow registration methods that exploit the EO image data or the 3D lidar data, or both, when the situation requires it. If the EO images contain few features or smooth areas, it is difficult to register them and compute 3D information using machine vision approaches. If the overlapping 3D points alone are used to register the 3D surfaces, the well-known Iterated Closest Point (ICP) algorithms may diverge due to outliers and noise.[14, 21] The approach taken in this paper is to use image features as a starting point. This is given in Algorithm 1, and is a modification of methods developed by Nelson and Boldt.[22, 23]

3.1 Detecting Harris Features

The first step in the basic algorithm is to detect Harris features[24] in each of the EO images. These are points that occur at corners and are detected as points that have large curvatures in both principal directions. Since the EO images are color, each of the color planes are processed individually, and the detected points from each of the planes are added to a list of feature points. Examples of feature points found in a pair of overlapping EO images is given in Figure 3.

3.2 Finding Putative Correspondences

The next step of the process is to find features in each EO image that possibly correspond to the same point in the 3D field of view. An assumption is made that the two views overlap and that the corresponding feature points have a high image correlation.

<div align="center">(a)   (b)</div>

Figure 4: Examples of putative correspondences found in the (a) first and (b) second images and marked with corresponding colored dots.

A window is centered on each feature, and the correlation between each of the features in the first image and each of the features in the second image is computed using

$$\gamma_{i,j} = \frac{\sum_{k=-N}^{N} \sum_{l=-N}^{N} \boldsymbol{u}_i(k,l) \boldsymbol{u'}_j(k,l)}{\left[\sum_{k=-N}^{N} \sum_{l=-N}^{N} \boldsymbol{u}_i(k,l)\right]\left[\sum_{k=-N}^{N} \sum_{l=-N}^{N} \boldsymbol{u'}_j(k,l)\right]}, \tag{1}$$

where the window size is $2N + 1$, (k, l) index points in the windows, and $\boldsymbol{u}_i$ and $\boldsymbol{u'}_j$ are the ith and jth feature points in the respective images. if $\gamma_{i,j}$ exceeds a threshold, the features are labeled as putative correspondences. An example of the resulting matches is given in Figure 4.

3.3 Random Sampling Consensus (RANSAC)

Once the putative correspondences have been determined, it is necessary to eliminate as many incorrect correspondences (outliers) as possible. A well-known method for eliminating outliers is to use the RANSAC algorithm.[25] RANSAC requires that a model must be used to separate inliers from outliers. This can be done by using a model that imposes a second constraint on the correspondences: they must satisfy the well-known epipolar constraint of projective geometry.[12] For this constraint to hold, the EO images must be calibrated to remove nonlinear distortion.[17]

A geometric interpretation of the constraint is illustrated in Figure 5. Any 3D point $\boldsymbol{x}$ visible in two images will project a point onto the image planes (given by $\boldsymbol{u}$ and $\boldsymbol{u'}$ in the figure). The plane formed by $\boldsymbol{x}$ and the COPs of the two cameras, $\boldsymbol{C}$ and $\boldsymbol{C'}$ (the epipolar plane), must contain the projected points. Therefore, if a putative correspondence does not meet this constraint, it is not a correct correspondence. More precisely, the test for the epipolar constraint is given by

$$\boldsymbol{u'}^T \boldsymbol{F} \boldsymbol{u} = 0, \tag{2}$$

where $\boldsymbol{F}$ is the fundamental matrix,[12] and $\boldsymbol{u}$ and $\boldsymbol{u'}$ are in homogeneous coordinates. The fundamental matrix can be computed from as few as eight correspondences, but is often computed from many more using least-squares (LS) estimation techniques.[26]

RANSAC is used as follows:

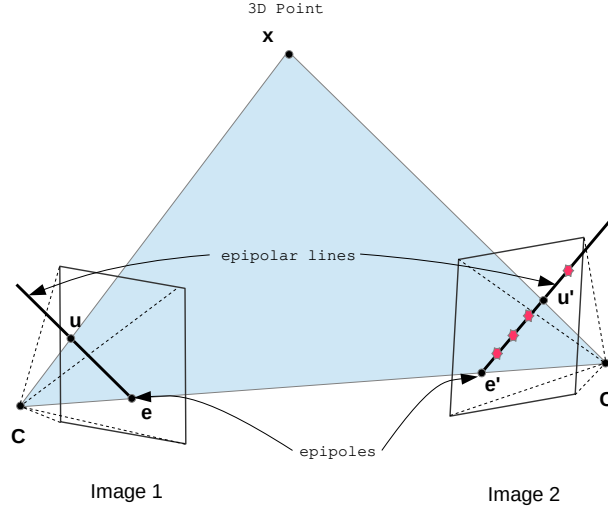1. Eight putative correspondences are selected randomly and the fundamental matrix $\boldsymbol{F}$ is computed.

Figure 5: Epipolar constraint required by projective geometry. The points u and u' represent corresponding image points projected from the 3D point x, and the red star points are erroneous correspondences to u satisfying the epipolar constraint. The COP of each camera is shown as C and C', respectively.

2. Each of the putative correspondences are tested by using (2), and if the product is below a threshold close to zero, the correspondence is counted.

3. Steps 1 and 2 are repeated for a desired number of iterations, and the set of correspondences with the largest count is taken to be the set of correct inliers.

3.4 Optimal Fundamental Matrix Computation and Testing

Once the largest set of inliers has been found, one more test of the complete set of putative correspondences is performed. First, the set of inliers from RANSAC is used to compute the optimal LS estimate of the fundamental matrix. This is then used in a final test to select the putative correspondences that meet both the correlation threshold and the epipolar constraint.

An example of the results of RANSAC and the final test is given in Figure 6. The white lines in the figure are the epipolar lines that each inlier corresponds to. Note that the number of correspondences is reduced to ones that meet the correlation and epipolar constraint conditions in the two EO images.

3.5 Point Cloud Matching using Fused EO Image/Lidar Data

At this point is is possible to use the fused EO image and lidar data to compute the transformation from the second camera reference frame to the first reference frame. No disparity computation is required. Since each EO correspondence found previously has a 3D point assigned to it (because it is a texel image), the 3D points can then be used to find the rigid body transformation from on lidar point cloud to the other.[27] The resulting transformation is found to be (R, t), where R is the rotation matrix and t is the translation vector from the second camera coordinate system to the first camera coordinate system.

4. ITERATIVE IMPROVEMENT IN REGISTRATION

Unfortunately, the epipolar constraint is necessary, but not sufficient for corresponding points, since any point on an epipolar line (where the epipolar plane intersects the image planes) will satisfy the constraint. This is illustrated in Figure 5, where erroneous correspondences that meet the epipolar constraint are marked as red stars. Thus, it is possible that erroneous correspondences can be found if EO points in one image correlate highly with points in the other image and lie on the same epiline that the correct correspondence lies on.
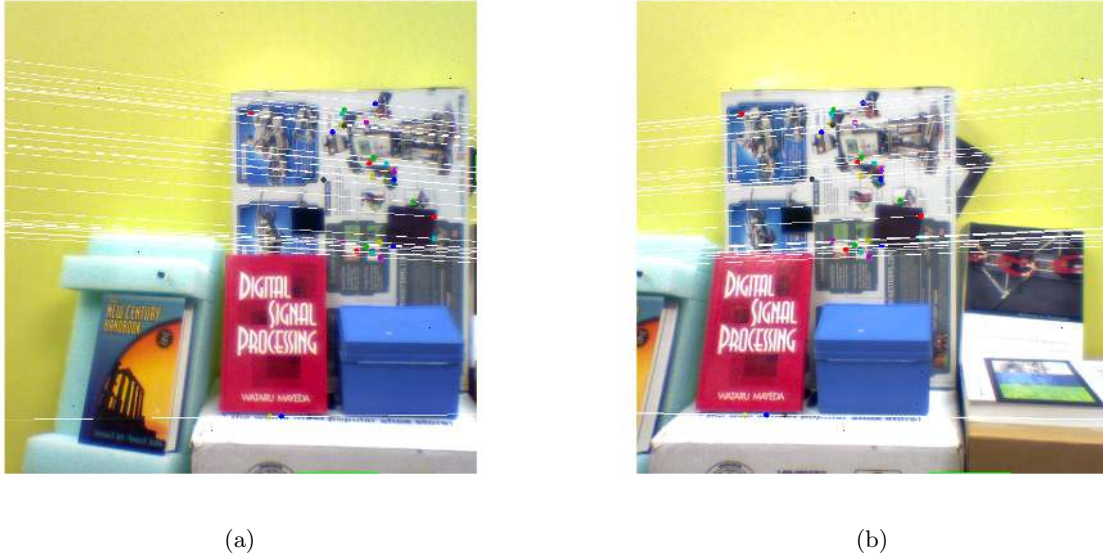
Figure 6: Examples of putative correspondences retained by RANSAC in the (a) first and (b) second images and marked with corresponding colored dots. The white lines are the epilines.

A solution to the problem is to guide the selection of putative correspondences by using the results of the first attempt at registration to eliminate correspondences that are in error. This leads to an iterative solution: if the first iteration of the registration process described in Section 3 is reasonably close, it can be used to guide a better selection of correspondences, which can then be used to estimate the 3D rigid transformation again.

Given an estimate of the transformation, $(\hat{\boldsymbol{R}}, \hat{\boldsymbol{t}})$, it is possible to test if putative correspondences meet the requirements of projective geometry. The epipolar test in Section 3.4 results in a set of correspondences that may be many-to-one for each feature point in the first EO image. It is necessary to prune the set to a one-to-one correspondence with the correct feature point, or even remove correspondences that don't fit the requirements of projective geometry. Therefore, each 3D point $\boldsymbol{u}_i$ in the first image is transformed into the second image coordinate frame using $(\hat{\boldsymbol{R}}, \hat{\boldsymbol{t}})$ and projected into the second EO image to create $\hat{\boldsymbol{u}}'_i$. If the correspondence is correct, $\hat{\boldsymbol{u}}'_i$ should be close to a feature point in the set of points for $\boldsymbol{u}'_i$. Two tests can then be made:

1. The feature point out of the set of possible correspondences that is closest to $\hat{\boldsymbol{u}}'_i$ is selected.

2. The 2D distance between the selected feature point $\boldsymbol{u}'_i$ and $\hat{\boldsymbol{u}}'_i$ must be below a threshold.

If none of the points in the set of $\boldsymbol{u}'_i$ pass the tests, the $(\boldsymbol{u}_i, \boldsymbol{u}'_i)$ correspondence is removed.

5. REGISTRATION RESULTS

The iterative technique of Section 4 was applied to pairs of texel images acquired from different positions and poses. A visual example of the results obtained are illustrated in Figure 7. In the figure, the top row ((a), (b), and (c)) shows the two texel images and the resulting registered texel image, and the bottom row ((d), (e), and (f)) shows the same images from a different perspective, revealing the 3D nature of the images. As can be seen in the figure, the registration of the two images is very good. The "seams" between the images occur next to the left textbook, and on the right edge of the blue box.

One of the most difficult registration problems is the case where two flat planes need to be registered. If 3D-only registration methods such as ICP are used, the registration process will often not converge, or converge to the wrong solution. In the noiseless case, any shift of the overlapping points will produce a minimum RMS registration error. For image-only registration and stereo 3D reconstruction, it is possible that multiple points may satisfy (to within the threshold) the epipolar constraint as described in Section 4, resulting in erroneous
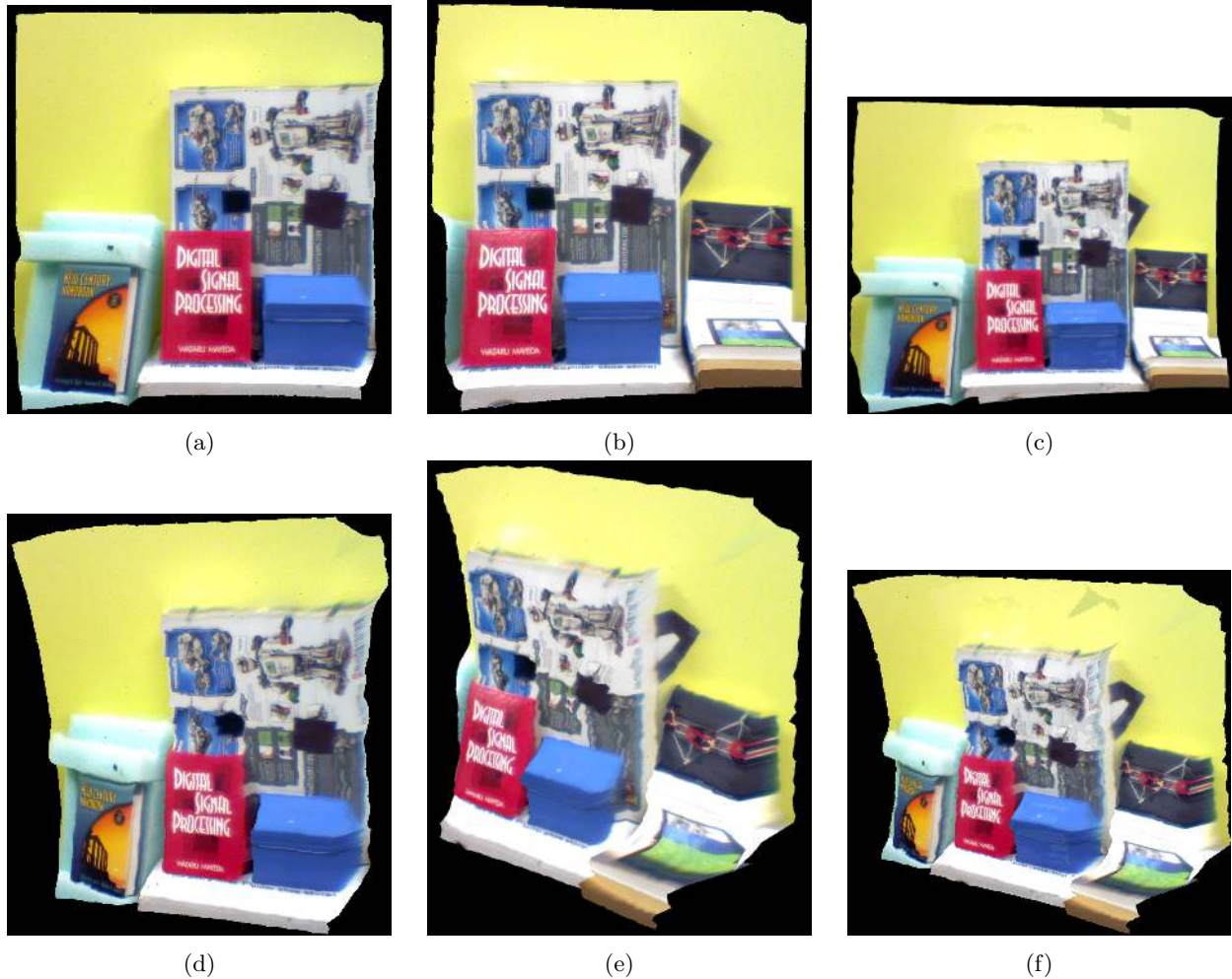
Figure 7: Registered texel images. (a) First texel image, (b) Second texel image, (c) Registered texel image, (d) First image from a different perspective, (e) Second image from a different perspective, (f) Registered images from a different perspective. All images are taken from a 3D viewing program.

correspondences. It is also difficult to compute stereo disparity if the images have large areas of constant color. An example of these problems is given in Figure 8, where (a) and (b) are from texel images of a flat wall taken with a right shift in perspective. Note that several correspondences are incorrectly declared because multiple corners of the grey bars are highly correlated and lie close to the same epiline. Registration using these correspondences leads to an evident mismatch, as observed in the registration of the grey bar chart and the broken edges of the black squares on the right.

An example of the improvement in putative point selection using point projection is given in Figure 9. The erroneous correspondences have been removed, leaving a set of correct correspondences for computing a new transformation $(\hat{\boldsymbol{R}}, \hat{\boldsymbol{t}})$, which is then used to register the two texel images.

The advantage of the proposed method over ICP is again illustrated with Figure 10. This figure shows the registration achieved using ICP with point-to-plane matching.[15] The registration is poor, with the Red textbook aligned with the blue box. The difficulty with this approach is that ICP tries to find the LS registration estimate using all of the points presented to it; satisfactory results require some preprocessing or prior knowledge to determine which points in the two point-clouds overlap.
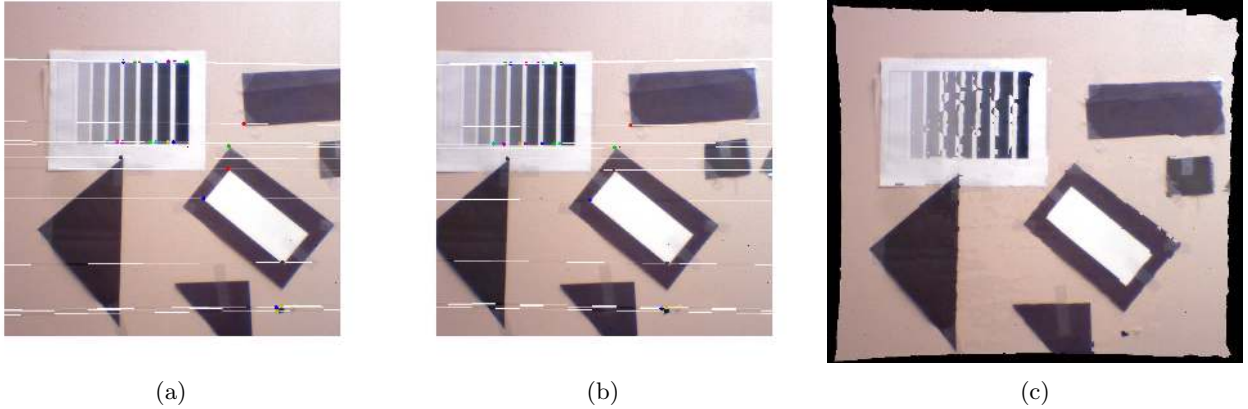
(a)   (b)   (c)

Figure 8: Examples of erroneous correspondences passing the epipolar constraint and the resulting registered texel image. (a) First EO image, (b) Second EO image, (c) Registered texel image. Correspondences are marked with the same color dot.
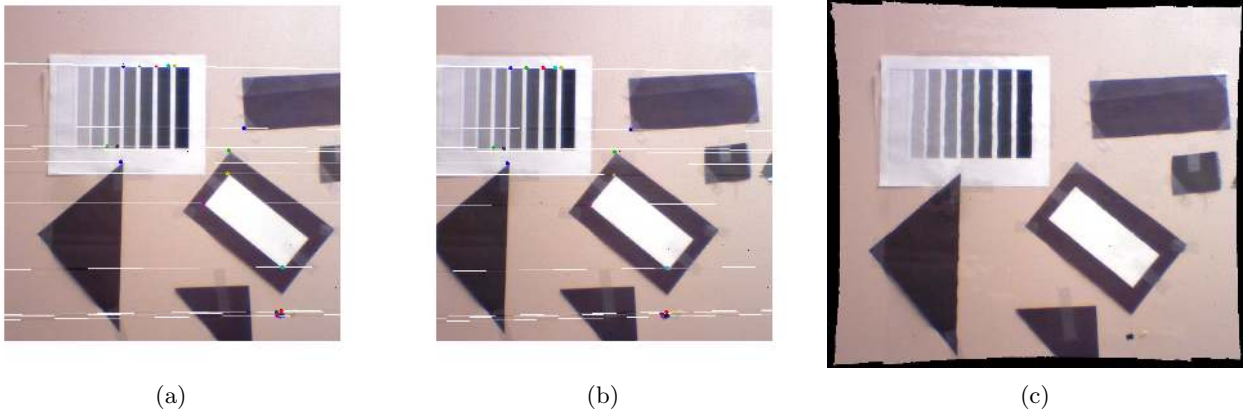

(a)   (b)   (c)

Figure 9: Improved correspondences after using 3D projection and the resulting registered texel image. (a) first EO image, (b) second EO image, (c) registered texel image.

Figure 10: Registration result achieved using ICP on the lidar point cloud only.

6. CONCLUSION

The method described in this paper has shown to perform very well in registering textured 3D point clouds (texel images) to create a 3D mosaic. If multiple texel images from different perspectives surrounding a scene are used,

it is possible to create a true 3D image of the scene.

Registration is based on the unique properties of texel images. Since the EO image is fused to the 3D lidar measurements at the sensor level, no 2D-to-3D registration, and resulting additional error, is introduced into the process. In addition, the registration of the texel images uses both the EO image and the 3D measurements. This allows for iterative improvement of the registration because the images must satisfy both the properties of projective geometry and 3D rigid transformation. Also, since measured 3D information is used in the registration, this method does not require decomposition of the fundamental matrix to find the 3D transformation, and resolution of the scale ambiguity often faced in 2D image-only methods.

There are many new and exciting areas for ongoing research in texel image registration. For example, the robustness of the techniques presented here must be studied further. The fused-dataset nature of texel images also open the possibility of image and shape superresolution techniques to improve both the resolution of the EO textures and the density and accuracy of the 3D data upon which the textures are applied.

REFERENCES

[1] Elstrom, M. D. and Smith, P. W., "Stereo-based registration of multi-sensor imagery for enhanced visualization of remote environments," in [*Proc. IEEE Int. Conf. Robotics and Automation*], **3**, 1948–1953, IEEE (May 1999).

[2] Hahne, U. and Alexa, M., "Combining Time-Of-Flight depth and stereo images without accurate extrinsic calibration," *Int. J. Intelligent Systems Technologies and Applications* **5**(3/4), 325–333 (2008).

[3] Zhang, J., Wang, L.-H., Li, D.-X., and Zhang, M., "High quality depth maps from stereo matching and ToF camera," in [*Soft Computing and Pattern Recognition (SoCPaR), 2011 International Conference of*], 68–72 (Oct. 2011).

[4] Stamos, I. and Allen, P. K., "Integration of range and image sensing for photorealistic 3D modeling," in [*Proc. IEEE Int. Conf. on Robotics and Automation*], 1435–1440, IEEE (Apr. 2000).

[5] Kurazume, R., Nishino, K., Zhang, Z., and Ikeuchi, K., "Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute," in [*The 5th Asian Conf. on Computer Vision*], 454–461 (Jan. 2002).

[6] Dias, P., Sequeira, V., Gonalves, J. G. M., and Vaz, F., "Automatic registration of laser reflectance and colour intensity images for 3d reconstruction," *Robotics and Autonomous Systems* **39**(3-4), 157–168 (2002).

[7] Deng, F., Li, S., Zhang, J., and Su, G., "Registration of LIDAR and optical images using multiple geometric features," in [*Proc. SPIE*], Maître, H., Liu, H. S. J., and Song, E., eds., *MIPPR 2007: Multispectral Image Processing* **6787**, 678714, SPIE, Wuhan, China (Nov. 2007).

[8] Borgeat, L., Poirier, G., Beraldin, A., Godin, G., Massicotte, P., and Picard, M., "A framework for the registration of color images with 3D models," in [*Proc. IEEE Int. Conf. Image Processing (ICIP)*], 69–72, IEEE (2009).

[9] Umeda, K., Godin, G., and Rioux, M., "Registration of range and color images using gradient constraints and range intensity images," in [*Proc. Int. Conf. Pattern Recoqition*], **3**, 12–15 (Aug. 2004).

[10] Deng, F., Li, S., and Su, G., "Mutual information based registration of LIDAR and optical images," in [*Proc. SPIE*], Ju, W. and Zhao, S., eds., *Geoinformatics 2007: Remotely Sensed Data and Information* **6752**, 675247, SPIE, Nanjing, China (May 2007).

[11] Scharstein, D. and Szeliski, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. of Comp. Vision* **47**, 7–42 (Apr. 2002).

[12] Hartley, R. I. and Zisserman, A., [*Multiple View Geometry in Computer Vision*], Cambridge University Press, second ed. (2003).

[13] Remondino, F., El-Hakim, S. F., Grün, A., and Zhang, L., "Turning images into 3-D models," *IEEE Signal Process. Mag.* , 55–64 (July 2008).

[14] Besl, P. J. and McKay, N. D., "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 239–256 (Feb. 1992).

[15] Rusinkiewicz, S. and Levoy, M., "Efficient variants of the ICP algorithm," in [*Third International Conference on 3-D Digital Imaging and Modeling*], 145–152 (2001).

[16] Grün, A. and Akca, D., "Least squares 3D surface matching," in [*ASPRS 2005 Annual Conference*], (on CDROM), ASPRS, Baltimore, MD (Mar. 2005).

[17] Budge, S. E., "A calibration-and-error correction method for improved texel (fused ladar/digital camera) images," in [*Laser Radar Technology and Applications XVII*], Turner, M. D. and Kamerman, G. W., eds., **8379**, 83790D, SPIE, Baltimore, Maryland, USA (Apr. 2012).

[18] Pack, R. T., Israelsen, P., and Sealy, K., "A co-boresighted synchronized ladar/EO imager for creating 3D images of dynamic scenes," in [*Laser Radar Technology and Applications X*], Kamerman, G. W., ed., **5791**, 42–50, SPIE (May 2005).

[19] Boldt, B. M., Budge, S. E., Pack, R. T., and Israelsen, P. D., "A handheld texel camera for acquiring near-instantaneous 3D images," in [*Proc. Asilomar Conf. Signals, Systems, and Computers*], (Nov. 2007).

[20] Pack, R. T., Swasey, J. A., Fullmer, R. R., Budge, S. E., Israelsen, P. D., Petersen, B., and Cook, T. D., "Eyesafe LADAR test-bed with coaxial color imager," in [*Laser Radar Technology and Applications XIV*], Turner, M. D. and Kamerman, G. W., eds., **7323**, 732303, SPIE, Orlando, FL, USA (Apr. 2009).

[21] Nishino, K. and Ikeuchi, K., "Robust simultaneous registration of multiple range images," in [*Proc. 5th Asian Conf. Computer Vision*], 454–461 (2002).

[22] Nelson, J. B., *Image-Based Correction of LADAR Pointing Estimates to Improve Merging of Multiple LADAR Point-Clouds*, Master's thesis, Utah State University, Logan, Utah (2006).

[23] Boldt, B. M., *Point Cloud Matching with a Handheld Texel Camera*, Master's thesis, Utah State University, Logan, Utah (2007).

[24] Harris, C. and Stephens, M., "A combined corner and edge detector," in [*In Proc. of Fourth Alvey Vision Conference*], 147–151 (1988).

[25] Fischler, M. A. and Bolles, R. C., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM* **24**, 381–395 (June 1981).

[26] Hartley, R. I., "In defense of the eight-point algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 580–593 (June 1997).

[27] Eggert, D. W., Lorusso, A., and Fisher, R. B., "Estimating 3-D rigid body transformations: a comparison of four major algorithms," *Machine Vision and Applications* **9**(5/6), 272–290 (1997).