Camera and LIDAR Fusion for Mapping of Actively Illuminated Subterranean Voids

Uland Wong[1], Ben Garney[2], Warren Whittaker[1] and Red Whittaker[1]

[1] Robotics Institute, Carnegie Mellon University
uyw@andrew.cmu.edu, warrenw@andrew.cmu.edu, red@cmu.edu

[2] PushButton Labs
Ben.Garney@gmail.com

Abstract A method is developed that improves the accuracy of super-resolution range maps over interpolation by fusing actively illuminated HDR camera imagery with LIDAR data in dark subterranean environments. The key approach is shape recovery from estimation of the illumination function and integration in a Markov Random Field (MRF) framework. A virtual reconstruction using data collected from the Bruceton Research Mine is presented.

1 Introduction

Mine accidents including those at Quecreek, Sago and Crandall Canyon highlight the urgency of estimating accurate 3D geometry in mines. Systems have been employed to map mines, from virtual reality systems for training rescue personnel [1] to automated survey robots and post accident investigation [2]. While many of these systems use state-of-the-art direct range measurement sensors, LIDAR sensors alone cannot meet the resolution, size, power or speed requirements to produce quality mine maps in a practical amount of time.

This research combines absolute range sensor data with high-resolution CCD imagery in a novel manner to achieve a quantitative increase in range data accuracy and density. In particular, the method targets application in artificial subterranean voids where assumptions can be used to constrain the image formulation problem. As both color and geometric information are of interest, cameras and range sensors commonly exist on modeling platforms [2]. Integration of the method presented here requires only calibration and low processing overhead.

The results from field experimentation in a working mine are discussed in detail. A dense visualization technique enabling mesh quality models to be displayed and updated in real-time on GPU hardware is explored. Lastly, a generalization of the method to similar domains in field robotics is made.

2 Prior Work

The fusion of range and imaging sensors to improve 3D model quality has been studied in depth [3,4,5,6]. A general model for fusing raw LIDAR and image data into super-resolution range images using a Markov Random Field (MRF) was explored in Diebel and Thrun's seminal paper [4]. MRFs are undirected graphs that represent dependencies between random variables and have been used extensively in computer vision for noise removal, feature matching, segmentation and inpainting (see [3]). The popularity of the MRF stems from the ability to model complex processes using only a specification of local interactions, the regular grid nature of CCD images and the maximum *a posteriori* (MAP) solution requiring only direct convex optimization in many cases.

Diebel and Thrun surmised that higher resolution intensity (color) data could be used to texture range images and increase the range accuracy of interpolated points. The results in a uniformly and sufficiently illuminated regular office environment are quite compelling. Cameras are able to turn LIDAR scans into dense range images with very low computational overhead. However, the assumption that an image provides relative range information, even locally, is tenuous in unstructured environments. Generating 3D geometry from a general 2D projection is an ill-posed problem. The ability of Diebel's method to smooth point clouds using areas of flat image information was convincingly shown, but the converse of enhancing a point cloud using image texture was not. Recent research in range/camera fusion using MRFs include [5,6]; all of which also target indoor application.

This research extends MRF-based super-resolution to subterranean environments such as mines, caves, lava tubes and sanitary pipes. These environments have unknown but slowly varying albedos with a dominant diffuse reflectance term. These naturally-dark, enclosed spaces also require active illumination to image, enabling the use of calibrated lighting. With these assumptions we are able to provide a stronger depth estimate for texturing the interpolated LIDAR data.

3 Markov Random Field Framework

A range image is used as the common representation for fusion. The 3D range cloud data is registered to the pinhole of the camera, forming a range map (R) via projection of distances onto the $n \times m$ image plane at equivalent resolution. Many pixels in the range map will not contain range measurements; these holes are filled from nearby data through bilinear or nearest neighbor interpolation. The color image data can be then converted to intensity values or used as a raw RGB vector (I). A lattice MRF is formed where there is a single range and intensity measurement associated with each node. We propose an MRF fusion method similar to that documented in [4] that numerically integrates the image gradient.

The range map potential (3.1) promotes agreement between the estimated variables and the interpolated range data. The smoothness prior (3.2) regularizes large changes in the range estimate and like the image potential (3.3) connects potential transfer from a node to its neighbors.

$$\Psi = w_1 \sum_{i \in L} \left(R_i - x_i \right) \tag{3.1}$$

$$\Omega = \beta \sum_{i \in L} \sum_{j \in N(i)} (x_j - x_i)^2 \tag{3.2}$$

$$\Phi = \alpha \sum_{i \in L} \sum_{j \in N(i)} \left(x_j + \nabla I_{ij} - x_i \right)^2 \tag{3.3}$$

$$\begin{aligned} \alpha &= w_2 \exp\left(-c \cdot \sigma\right) \\ \beta &= w_3 \left(1 - \exp\left(-c \cdot \sigma\right)\right) \end{aligned} \tag{3.4}$$

The image gradient is a reasonable predictor of depth change across neighboring pixels. However, integrating the gradient to produce depths over a large locality is prone to drastic shape distortions. The range estimate can be used to regularize numerical integration of the intensity gradient. The weights α and β are relatively scaled by an interpolation distance uncertainty (σ) for some weights w_2 and w_3 (3.4). σ can be generated from the range image during inpainting by using the Matlab command BWDIST, for example.

$$p(x \mid R, I, \sigma) = \tfrac{1}{Z} \exp\left(-\tfrac{1}{2}\left(\Psi + \Omega + \Phi\right)\right) \tag{3.5}$$

$$x_{MAP} = \arg\min_x f(\Psi + \Phi + \Omega) \tag{3.6}$$

Solving for the MAP of the distribution requires running a gradient descent algorithm on the target variables x in 3.5-3.6, where Z is the partition function [4].

4 Structure from Shading

The image gradient ∇I_{ij} in (3.3) can apply to either raw pixel data or better estimates of depth from the camera. As scene geometry cannot be ascertained from a single image without assumptions, often no better estimate exists. Definite reconstruction requires knowledge of image formation parameters like light field, surface reflectance (BRDF) and albedos. However, if assumptions like those commonly made in Shape-from-Shading are valid, the number of unknowns is greatly reduced.

The illumination and reflectance assumptions are appropriate for subterranean environments. Most dry underground mines and caves are located in Lambertian rock and many coal mine interiors are additionally covered with diffuse material

like Shotcrete [7]. Low amounts of metallic meshing, industrial equipment, water and retro-reflectors are present, but the contribution of these specular surfaces can be reduced using the method documented below and in [10]. Robots in these naturally dark environments can be fitted to carry small area light sources for photography which produce simple light fields.

The MRF image observation (I) is estimated using Shape-from-Shading given the above assumptions. A lightness-based direct normal estimation method which uses range information is given below, but other techniques exist [8,9]. This method factors range information to allow varying albedos and trades accuracy for feature preservation. The effect of the light source's irradiance fall-off is first removed from the raw image data (E_0). We assume the following irradiance correction model for small area sources (4.1):

$$E_{unbiased} = \gamma(E_0) \cdot R^n \tag{4.1}$$

The radiometric function (γ) maps pixel values to irradiance, (R) is the interpolated depth estimate and (n) is the irradiance fall-off factor. For ideal point sources $n = 2.0$, while $n < 2.0$ for near-field area sources. The experimental setup described below exhibits an empirical decay of $n = 1.265$. The corrected image (E_c) is devoid of a near-field illumination intensity bias.

Converting RGB color into a single intensity value provides compactness and symmetry, and also minimizes chromaticity effects. Color space transformations such as CieLAB or YCbCr are often used to heuristically isolate the lightness component of an image, discarding chromaticity and albedo. Zickler's SUV transformation [10] describes a class of physics-based specular-invariant color spaces produced by rotating the RGB space such that a single channel is aligned with the illuminant color vectors. This method has produced excellent results with single-source images and enables many Lambertian algorithms to handle a large set of environments with specularities. The specular invariant image, as defined in eqs. 4.2-4.3, is used in experimentation:

$$[s,u,v]^T = R_r(\theta) \cdot \left[E^{(r)}_{unbiased}, E^{(g)}_{unbiased}, E^{(b)}_{unbiased} \right]^T \tag{4.2}$$

$$E_{inv} = \sqrt{u^2 + v^2} \tag{4.3}$$

$R_r(\theta)$ is defined as a (3×3) rotation matrix that aligns the red channel of an $\{r,g,b\}$ triple with the source color. The magnitude of the $\{u,v\}$ components is taken to be the diffuse image.

An albedo map is subsequently generated from the diffuse image using Blake's method for lightness computation [11]. Perceived intensity is a multiplicative relationship between surface slant angle and reflectance. The log image separates these components into additive terms. Scene albedos can be recovered from the gradient of the log diffuse image by thresholding to remove small changes and in-

tegrating. It is noted that the problem can be recast as finding the log albedo map (δ) that minimizes the following [11]:

$$\arg\min_{\delta} \left| \frac{\partial}{\partial x}\delta - T_{\sigma}\left(\frac{\partial}{\partial x}\log E_{inv}\right)\right|^2 + \left| \frac{\partial}{\partial y}\delta - T_{\sigma}\left(\frac{\partial}{\partial y}\log E_{inv}\right)\right|^2 \tag{4.4}$$

where (T_{σ}) is the threshold function. Exponentiating (δ) with the proper constant of integration produces the albedo values (4.5). The constant can be estimated from the range data to minimize depth discrepancy in the reconstruction.

$$\rho_{est} = \exp(\delta + c) \tag{4.5}$$

$$E_{inv} = \rho |n||l| \cos(\theta_{nl}) \tag{4.6}$$

$$\theta_{nl} = \arccos\left(\frac{E_{inv}}{\rho_{est}}\right) \tag{4.7}$$

The polar estimates (θ_{nl}) are combined with azimuth estimates (ϕ) from the range image and converted to gradients for integration in the MRF.

5 Experimental Results

The experimental setup uses both a continuously rotating planar LIDAR scanner and an 8 megapixel DSLR camera mounted to a mine robot. A small area light source is also mounted along the same axis to minimize cast shadows in the image. This replaces the normal flood lighting for the imager. The scanner has a practical throughput of ~40,000 points per second. The points are aligned along concentric rings with 0.5° angular separation in a 180° hemisphere in front of the unit. The camera takes hemispherical images using a constant angular resolution fisheye lens with a 182° field of view. The sensor mounting configuration and example data are shown in Fig. 1 below.

Fig. 1. (Left) Experimental setup with 1. LIDAR scanner. 2. Fisheye Camera, 3. Light Source. (Center) Raw fisheye imagery. (Right) Ground truth range image.

Thirty complete datasets consisting of LIDAR scans, High Dynamic Range (HDR) imagery and robot odometry were collected from the Bruceton Research

Coal Mine in Pittsburgh, PA. LIDAR scans averaged 600,000 points. HDR images were each generated from a series of 5 images corresponding to exposures times of {¼, ½, 1, 2, 4} seconds using the method described in [12]. The 1.0 second exposure image was used as the Low Dynamic Range (LDR) reference image for analysis. An additional 16 datasets of LDR-only imagery were also collected.

A ground truth range map was generated for each LIDAR scan using the full point cloud. Multiple measurements mapping to the same pixel were averaged. The scans were subsequently down-sampled to 25,000 points and interpolated into a range image for testing the method. The datasets were further partitioned into 25 test sets and 5 training sets. Optimal weighting factors were learned using a simplex search on the training set, while validation occurred in the test set.

The proposed method was compared against Diebel's method and raw interpolation. The mean per-pixel L_1 norm (Manhattan distance) between the reconstructed range map and the ground truth map was used as a benchmark for comparison. Ground truth data points outside the convex hull of LIDAR values in the interpolated map were discarded due to skew in scoring extrapolated points. The usable pixel area is determined for each scan by the number of saturated pixels, the range image convex hull and removal of high-gradient probable error values.

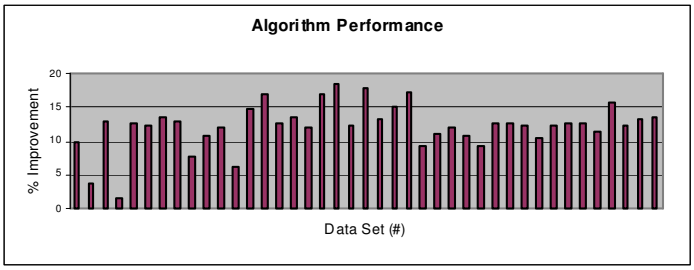

Fig. 2. Reconstruction Improvement vs. Raw Interpolation.

The results of the experiment are summarized in Table 1 and Fig. 2. Fig. 3 and Fig. 4 show an example reconstruction from a single view point. The scene features a yellow nylon mine curtain on the left side, wooden cribbing stacks on the right and aluminum meshing integrated into a mostly exposed ceiling.

Table 1. Summary of Results

Quantity	Details
Total Test Datasets	41
{HDR, LDR-only} Datasets	{25, 16}
Interpolation Improvement	
Mean	12.2%
Max, Min	19.2%, 3%
Density Statistics	

LIDAR downsample	25,000 points
Ground Truth LIDAR	669,834 points
Mean Resultant	1,045,358 points
Mean Increase	41.8 x
Image Usability Information	
LDR Saturated	3.17% of total pixels
HDR Saturated	4.20×10^{-2} % of pixels
HDR Accuracy Increase	20.5% over LDR-only
HDR Density Increase	51.5% over LDR-only

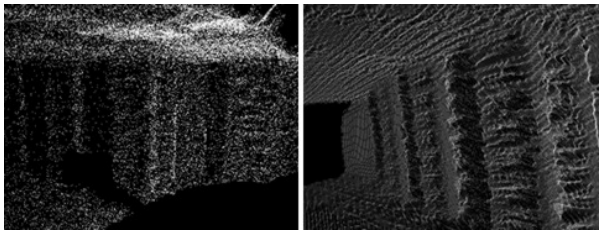

Fig. 3. Point Cloud of Cribbing. Low resolution cloud (left) and high resolution reconstruction from algorithm (right) showing stacked timbers supporting the roof.

Fig. 4. Colorized 3D Reconstruction. Full scene (left) and mine curtain detail (inset and right).

Additional data of two corridors were also collected at the Bruceton Mine along evenly spaced intervals roughly 3 meters apart. Using robot odometry and Iterative Closest Point (ICP) alignment, multiple scans were up-sampled using the proposed technique, fused together and color/illumination compensated. These models represent some of densest, most comprehensive mine reconstructions to date using a mobile robot. The results appear below:

Table 2. Corridor Modeling Statistics.

Model #	# of Scans	# of Images	# Points
1	4	16	5,543,451
2	8	32	9,680,105

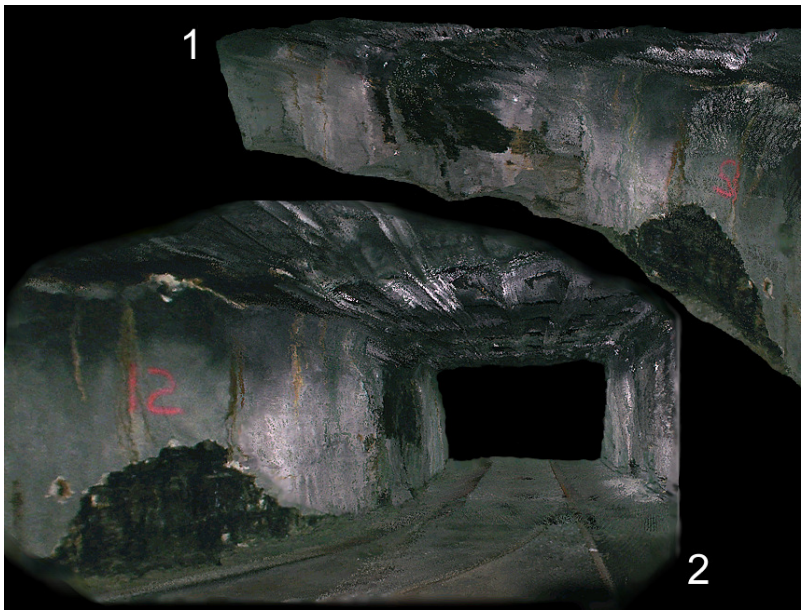

Fig. 5. Mine Corridor ICP model. (1) External view. (2) Internal view with rail tracks.

The results are displayed using a hole-filling method similar to the multi-scale push-pull technique in [13]. This display system is adapted to benefit from high density clouds generated using super-resolution methods. Point clouds are rendered with push-pull interpolation in image space. A min-depth check and kernel density estimator are used to resolve edge discontinuities and remove occluded background measurements. The utilization of texture in-painting for both color interpolation and depth reconstruction provides the viewer with graphical continuity as well as proper occlusions, which standard point displays lack. In addition to fast rendering of huge datasets, the renderer allows the model to be updated in real time as new data arrives without costly re-meshing operations. The system can generate real-time (>30Hz) imagery at 1080p HD resolution on commodity (GeForce GTX 260) hardware with point clouds of greater than 5 million points

7 Analysis

The results show that the method increases interpolation accuracy by up to 20% on the Bruceton Mine data, with an average improvement of 12%. The fisheye-spinner setup features density increases up to 70 fold, with an average of 40x increase in density (Table 1). Of note is that real resolution is created where LIDAR beam physics dictate a maximum angular resolution. This is apparent in 3D scan-

ning mechanisms that actuate a planar sensor, where an increase in data collection time results in diminishing resolution returns.

To validate that true information is being stored in the interpolated values, a sliding-window 15x15 pixel Pearson correlation was performed. As shown in Fig. 6, the shaded image provides significant information about the ground truth that is not contained in interpolation. The fused range map correlates more than either source individually, concurring with the error estimation benchmark. While Diebel's method shows a numerical increase in accuracy, it is not statistically significant. This is corroborated by almost equal amounts of strongly negative and positive correlation in the raw image data.

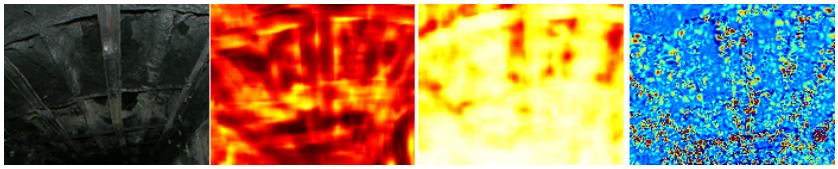

Fig. 6. (Left to right) Roof supports covered in Shotcrete, Image to ground truth correlation, Shaded image to ground truth correlation, and Reconstruction error reduction. Scale is brown to white over [-1, 1] for correlation and navy blue to red over [-0.025m, 0.05m] for error reduction.

The method encounters several drawbacks that prevent the fused result from achieving the same accuracy as LIDAR scans of equivalent density. Resulting range images are vulnerable to artifacts typical of raw interpolation, although to a lesser degree. Most reconstruction error occurs at occlusion edges where neighboring LIDAR points have large disparities. Regularization terms tend to over-smooth these edges and shading cues are ill-behaved due to cast shadows, among other reasons [9,11]. Attempting to isolate these specific edges in the image is difficult due to image noise, lighting and material specific effects and is not addressed in this research (see [5,14]).

8 Conclusion

A method was presented that fuses actively illuminated CCD imagery and LIDAR data. The method demonstrates increases in range accuracy of up to 20% on experimental data over interpolation and increases in measurement density of up to 70x using the experimental setup. The improvements are a result of calibrated imaging using additional knowledge of the image formulation model to reconstruct a 3D observation of the scene. This research demonstrated the efficacy of multi-sensor mapping systems as well as calibrated imaging for field robots.

Perhaps the greatest argument for range/image super-resolution is that it is easily bootstrapped to existing systems. Subterranean robots already require light sources for photography as well as range sensors for mapping and many high-

throughput commercial scanners feature co-located cameras. The general use of illumination information for super-resolution is also applicable to other domains in field robotics. Planetary robots are likely to encounter highly diffuse environments (i.e. Mars) or characterizable reflectances on bodies lacking scattering atmospheres (i.e. moon, asteroids). Such development is likely to increase the safety of exploration and prospecting on the moon, where sensing is secondary to payload and comes at a premium cost.

In future work, agile and high accuracy applications will benefit from one or more actuated sensors. With an actuated camera, the technique can be used to zoom in on regions of interest and selectively up-sample a scan. With an actuated range scanner, such as a focal plane array LIDAR with variable optics, a static camera may be used to reconstruct rough low frequency data from a preliminary scan and detect areas of high frequency, while a second LIDAR pass focuses specifically on these areas. This setup can drastically compress the amount of data taken and reduce time required while producing an optimal reconstruction given certain throughput constraints.

References

1. P. Boulanger, J.-F. Lapointe. Creation of a Live Virtual Reality Model of Mining Environments from Sensor Fusion. In Proc. Minespace 2001.
2. A. Morris, D. Ferguson, et al. Recent Developments in Subterranean Robotics. Journal of Field Robotics, Vol. 23, No. 1, Jan 2006, pp. 35-57.
3. S. Li. *Markov Random Field Modeling in Image Analysis*. 2001.
4. J. Diebel, S. Thrun. An Application of Markov Random Fields to Range Sensing. In NIPS 2005.
5. L. Torres-Mendez, G. Dudek. Inter-Image Statistics for 3D Environment Modeling. IJCV (2008) 79. pp 137-158.
6. S. Gould, P. Baumstarck, M. Quigley, et al. Integrating Visual and Range Data for Robotic Object Detection. In ECCV 2008.
7. M. Clements. Shotcreting in Australian Underground Mines: A Decade of Rapid Improvement. In Shotcrete Spring 2003.
8. R. Zhang, P. Tsai, J. Cryer, M. Shah. Shape from Shading: A Survey. IEEE PAMI Vol. 21, No. 8, August 1999.
9. A. Braquelaire., B. Kerautret. Reconstruction of Lambertian Surfaces by Discrete Equal Height Contours and Regions Propagation. Image and Vision Computing, 2005. Volume 23(2), pp. 177-189.
10. S. Mallick, T. Zickler, D. Kriegman, P. Belhumeur. Beyond Lambert: Reconstructing Specular Surfaces Using Color. CVPR 2005.
11. P. Worthington. Re-illuminating single images using Albedo estimation. In Pattern Recognition 2005.
12. P. Debevec, J. Malik. Recovering High Dynamic Range Radiance Maps from Photographs. In SIGGRAPH 1997.
13. J. Grossman, W. Dally. Point Sample Rendering. In Rendering Techniques 1998.
14. Q. Yang, R. Yang, J. Davis, et al. Spatial-Depth Super Resolution for Range images. In CVPR 2007.